METHODOLOGICAL DEVELOPMENT AND ADVANCES FOR JOINT MODELLING OF LONGITUDINAL AND TIME-TO-EVENT DATA

G BUYRUKOGLU

PhD

2018

METHODOLOGICAL DEVELOPMENT AND ADVANCES FOR JOINT MODELLING OF LONGITUDINAL AND TIME-TO-EVENT DATA

GONCA BUYRUKOGLU

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the Department of Mathematics, Physics and Electrical Engineering

July 2018

Abstract

Univariate joint modelling of longitudinal and time-to-event data is a simultaneous analysis of repeated measurements taken from the same individual over time, until an event of interest occurs. This method has attracted increasing interest in the literature over the last two decades. In practice, clinical studies are increasingly likely to record more complex data structures (such as multilevel longitudinal data or multiple longitudinal profiles, along with event time data) than single longitudinal and event time data. This thesis develops a methodology and software for both multilevel and multivariate joint models accounting for complex longitudinal data, by focusing on random effects selection models, where information from the longitudinal trajectories is used to inform the event-time process. The research also assesses the power of the score test, which is a prognostic tool to investigate the association between submodels, before fitting potentially complex and computationally intensive joint models under a variety of scenarios. The methodology is tested via simulation studies, and implemented in various real datasets. The results show that the advanced joint models can provide unbiased estimators when the model is specified correctly such that it utilizes all available data, and that the score test is a powerful tool when the longitudinal profile is highly associated with the event time data. Based on preliminary findings using discrimination measures, the advanced joint models should be preferred in case of complex longitudinal data in order to improve the predictive capability of the model.

Dedication

To mum and dad

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. Any ethical clearance for the research presented in this thesis (Research project RE-EE-14-150330-5519570105a64) has been approved. Approval has been sought and granted by the Faculty Ethics Committee on 30 March 2015.

I declare that the Word Count of this Thesis is 37460 words.

Name: Gonca Buyrukoglu

Signature:

Date:

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Pete Philipson and Dr. Nan Lin for the continuous support they provided during my PhD study and related research, and for their patience, motivation, and immense knowledge. Their guidance helped me during the entire research process and writing of this thesis. I could not have imagined having better supervisors and mentors for my PhD study.

I would like to express my sincere gratitude to the Republic of Turkey's Ministry of National Education for the funding provided for my research studentship and the warm support provided throughout.

Very special thanks go to my husband, Selim, my daughter, Feyza and my son, Yusuf for their love, support, understanding and the many sacrifices they made throughout the period of my study. Also, I offer my grateful thanks to my family and siblings for their ever-constant love and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The joint modelling of longitudinal and survival data has become a highly-active field of research over the past twenty years. The original methodology for a univariate random effects joint model has been developed by Wulfsohn and Tsiatis (1997) and extended by Henderson et al. (2000), and has become known as the standard joint model. Recent developments have considered joint models for multivariate longitudinal data and a single survival data (Lin et al., 2002), as well as for a single longitudinal and multivariate survival data (Williamson et al., 2008). The standard joint model considers a linear mixed effects submodel for the longitudinal data and a Cox-based submodel for the time-to-event data, and links these submodels through shared random effects. The various random effects' structure can change the nature of the association.

Historically, joint models were applied to data arising from clinical trials. More recently, trials across multiple centres and countries have been conducted to combat serious diseases. Furthermore, on a larger scale, there has been considerable interest in utilising electronic healthcare databases. Such databases can link repeated measurements with the event's history record. Each of these scenarios points to the need for joint modelling approaches that are considered to be the most efficient way to capture available information from two (or more) data sources, giving rise to longitudinal data and survival data. This chapter describes the fundamentals of longitudinal and survival data analysis.

1.1 Longitudinal data analysis

Longitudinal data is very common in practice, featuring in numerous applications such as epidemiology, clinical trials, economics, industry and biology. In longitudinal studies, individuals are followed over time and measurements are taken from the same individual repeatedly. The main goal of a longitudinal study is to characterize the change in response over time, and find out the factors that affect the change. Responses between subjects may be independent; however, repeated measurements taken from the same individual are very likely to be correlated. This causes the violation of the independence assumption, which is one of the fundamental assumptions in traditional statistical methodology. This dependency amongst measurements within individuals makes longitudinal data analysis a specialist statistical research area.

The mixed effects modelling framework (also known as the random effects models) is an important data analytic class for longitudinal data analysis. These models consider the relationship between serial observations on the same unit. While the general covariance structure is hardly applicable to these kinds of models, two-stage random effects models can be applied easily. In these models, the explanatory variables are obtained by regression parameters as in the traditional regression, where the random effects parameters vary across individuals and capture any subject-specific deviations. The subject-specific random effects parameters are specified at the second stage. A general family of such models is discussed in the seminal paper of Laird and Ware (1982), which incorporates both growth models and random effects models.

Further development of the statistical methodology for longitudinal data analysis continued over the following three decades. For instance, Diggle (1988) proposed a linear model for longitudinal data by accounting for the serial correlation structure within the same unit, and the utilization of the variogram of residuals to determine the suitable correlation structure of repeated measurements. On the other hand, Taylor et al. (1994) developed a plausible and

parsimonious model to describe the stochastic process underlying the patterns of longitudinal data, which enables one to understand if the subjects keep their trajectories. In addition, Jennrich and Schluchter (1986) presented a parameter estimation approach for unbalanced and incomplete repeated measurements using the maximum likelihood (ML) method along with some numerical algorithms such as Newton-Raphson methods and the expectation maximization (EM) algorithm, which was originally discussed by Dempster et al. (1977). In this context, Lindstrom and Bates (1988) improved the implementation of the algorithm presented by Jennrich and Schluchter (1986), and sped up the convergence at each iteration. Longitudinal data is discussed in more details in numerous literature sources, such as the books authored by Verbeke (1997), Diggle (2002), Fitzmaurice et al. (2008) and Fitzmaurice et al. (2012).

The assumption of the most well-known linear mixed effects model for a continuous response is that the random effects and the within-subject errors have normal distribution. This assumption makes the model sensitive to outliers, and these outliers can be problematic for mixed effects models, as they are likely to be seen in the random effects and make them difficult to detect in practice. As a solution, Pinheiro et al. (2001) proposed random effects models having a repeated measurement multivariate t distribution to be able to cope with this problematic issue.

Researchers use many different kinds of statistical software programmes to implement longitudinal data analysis. However, the R programming language, which is an open-source software programme, is preferred in this work, and is used throughout this thesis (R Core Team, 2017). There are two packages available to implement longitudinal data analysis in R, namely nlme (Pinheiro et al., 2017) and lme4 (Bates et al., 2015). Whilst these packages are able to fit the Laird-Ware model, nlme can also fit random effects models with a stationary Gaussian process (SGP). Due to the superiority of the nlme package and the chance of the possible extension to the incorporation of the SGP in joint model, it is preferred to fit the Laird-Ware model throughout this thesis. The R codes

regarding all data simulation, analysis of the simulated data and real data can be found at `https://github.com/goncabuyrukoglu`.

1.2 Survival data analysis

Survival data (also known as time-to-event data) are the main interest of many fields, such as epidemiology, biology and medicine, in which the primary interest lies in the time it takes from a given baseline for an event of interest to occur, and the identification of any factors related to this event. An example of this would be studying the effect of a treatment on the time-to-dementia (event) since diagnosis of mild cognitive impairment. The analysis of such data is described as time-to-event analysis or survival analysis. This section provides a brief introduction to the survival data analysis and an overview of the current literature.

The probability of experiencing an event of interest, the rate at which the event occurs, and how the rate changes among groups are some of the outcomes of interest from time-to-event analysis. A distinguishable characteristic of survival data is that, not all participants experience the event within the follow-up time. This leads to censoring problems. There are three types of censoring problems, namely right censoring, left censoring and interval censoring.

Right censoring happens when the event of interest does not appear within the specified time frame. It is the most common censoring type. For example, in a cohort study, patients who have been diagnosed with breast cancer are followed, and two groups of treatment are investigated in terms of the time-to-death, and are compared to each other. While, some patients may live up to 30 years after treatment, it is generally not feasible to continue to study after a specified point due to resources constraints, such as finance or time. If a patient has not experienced the event within the follow-up time, this is described as administrative censoring and considered as non-informative censoring. If a patient moves from one place to another or withdraws from the study for rea-

sons related to the outcome, then this is called informative censoring.

Left censoring occurs when the event of interest happens before a patient starts to be observed or before the outcome is verified. It generally occurs when the event of interest is the relapse of a disease. The data cannot be left censored when the event is death. If a patient already has cancer before the follow-up, it can be an example of left censored data.

Interval censoring occurs when the event of interest occurs within an interval and covers the other two scenarios (i.e. is a combination of left and right censoring). It occurs when the event time is between two time points. Following the previous example, if the relapse does not occur in the first visit of the clinic, and is instead found in the following visit, then the actual time to relapse is between the first visit and second visit.

Event times are necessarily positive and are typically right skewed data. Considering the censoring, right skewness of data, and having the interest mostly on hazard and survival functions makes survival data analysis complicated, requiring the development of bespoke statistical methods and software programmes, as in the case of longitudinal data analysis.

Survival data was first addressed by Kaplan and Meier (1958), in which survival data analysis methods compared survival curves amongst groups. A remarkable development in medical and statistical research was published by Sir David Cox in 1972 (Cox, 1972), which is one of the most important research publications in this field, with over 46,000 citations (Google Scholar, January 2018) since its publication. The method presented in the research is a semi parametric approach with the incorporation of explanatory variables, and has become known as the Cox proportional hazards model. The Cox model estimates the parameters via the partial likelihood method by generalizing the ideas of conditional and marginal likelihood (Cox, 1975), and does not assume any particular form of baseline hazard function. It is one of the main

approaches and is often used as the default choice in the analysis of survival data. The other types of approaches (which assume exponential or Weibull distribution) for the survival times are fully parametric approaches, such as that of Collett (2015). The original Cox model deals with only constant covariates through time. It was then extended to incorporate time-dependent covariates through a counting process approach (Andersen and Gill, 1982), based on the Aalen model BAŞAR (2017). Vaida and Xu (2000) proposed a general proportional hazards model with random effects (known as the frailty model), and Yamaguchi et al. (2002) extended it to allow for centre variation in the treatment effects, as well as baseline risks.

Oakes (2013) published a useful overview with some key ideas and models in survival analysis. However, the interested reader is referred to Kleinbaum (1998) as a starting self-learning resource on survival analysis, while for advanced methods, the reader may find it useful to be referred to the extended survival data analysis textbooks written by Kalbfleisch and Prentice (2011), Lawless (2011), Lee and Wang (2003) and Hosmer et al. (2008). Furthermore, Andersen et al. (1993) is an interesting resource for survival data analysis with counting processes. In addition, the `survival` package (Therneau, 2015) is available to implement survival data analysis, and contains some auxiliary functions such as the comparisons of survival curves, along with fitting of a range of models.

1.3 Joint modelling of longitudinal and survival data

In medical studies, longitudinal outcomes are often intrinsically collected with event time data, even if this is not ostensibly the focus of the study, as it is sometimes a natural requirement of the longitudinal data. However, these outcomes are analysed separately, despite the obvious potential that the two types of outcomes are associated. When this potential association is to be explored, a more complex method to analyse the two data types needs to be considered. This brings us to a popular, novel and rapidly growing research area over the

past twenty years, namely the field of joint modelling of longitudinal and survival data (also known as simultaneous analysis of the two types of outcomes).

The most well-known approach, also known as the standard joint modelling of longitudinal and time-to-event data, was developed by Wulfsohn and Tsiatis (1997), combining the linear mixed effect model with a proportional hazards model through shared random effects with maximum likelihood, implemented via the EM algorithm, which is used for parameters estimation. This method characterises the association between the two outcomes. A joint model allows much greater insight into both outcomes, with reduction of bias, making an optimal use of the available data with an attempt to disentangle the underlying association in an interpretable manner (Lawrence Gould et al., 2015). An early study to accommodate longitudinal data as a time dependent covariate into survival models was carried out by Andersen and Gill (1982). However, the drawback of this study is the assumption of perfectly measured covariates and their availability at each time point. In practice, it is nearly impossible for this assumption to hold in any moderate sized study. Real life observations are prone to measurement errors, they are taken possibly intermittently and of unequally spaced time points for the repeated measurements, and possibly subject to censoring event times. A further study to improve this method was proposed by Tsiatis et al. (1995), and is referred to as a two-stage analysis of longitudinal and time-to-event data. In the first stage, the method models the longitudinal data using a repeated measurements random components model (in a separate analysis), and in the second stage, the parameters in the survival model are estimated by considering the observed values as a time dependent covariate in the Cox model. Nevertheless, Sweeting and Thompson (2011) stated that a two-stage model can severely underestimate the association between the underlying longitudinal value and event hazard. Simultaneous analysis of the two outcomes handles the aforementioned drawbacks of the two-stage approach.

Berzuini and Larizza (1996) investigated the simultaneous modelling of longitudinal and survival data from the Bayesian perspective, and Faucett and

Thomas (1996) used MCMC techniques of Gibbs sampling to estimate the joint posterior distribution of the unknown parameters. These studies combine the random effects for longitudinal data and the proportional hazards model for event times data. Henderson et al. (2000) extended the joint modelling framework with a flexible parametrisation of association structure between the two outcomes, incorporated into special cases of the model, and described a Monte Carlo EM estimation procedure considering the two components to be linked through a latent stationary Gaussian process. While Henderson et al. (2000) used maximum likelihood estimation via the EM algorithm using antithetic pairs, Guo and Carlin (2004) investigated flexible parametrisation with a Bayesian paradigm.

As the joint modelling is an active and popular research field, there are now many published studies that have extended this by developing either the longitudinal or survival side of the modelling. Sousa (2011) reviewed the joint modelling of longitudinal and survival data in detail, and stated that different factorisations of the joint distribution can bring about different model interpretations such as pattern-mixture and selection models, which are explained in detail in the following chapter. The interested reader may be referred to read the review papers published by McCrink et al. (2013), Lawrence Gould et al. (2015) and Sousa (2011) for extensive information on joint models. For further reading, the book-length material of Elashoff et al. (2016) and Rizopoulos (2012), which are solely dedicated to joint models, can be useful. There are four packages for joint models in the `R` programming language, namely `JM` (Rizopoulos, 2010), `joineR` (Philipson et al., 2017) for ML estimation, `JMbayes` (Rizopoulos, 2016) for Bayesian approaches, and `joineRML` (Hickey et al., 2017) for joint models of multivariate longitudinal data and time-to-event outcomes. The `st jm` package is also available to fit shared parameter joint models in `Stata` using the maximum likelihood method.

1.3.1 Links between missing data mechanisms and joint modelling

Missing data is a common occurence and a major challenge in longitudinal studies. Longitudinal studies are designed to collect data on every single subject in sample; however it is often not the case encountered in practice. Some data are not complete due to a variety of reasons and this brings about the term 'missing data' (Little and Rubin, 2014; Molenberghs and Kenward, 2007; Ibrahim and Molenberghs, 2009). Rubin (1976) classified the missing data mechanisms into three categories:

- MCAR (missing completely at random): Missingness depends neither on observed nor unobserved measurements.
- MAR (missing at random): Missingness depends on observed measurements, but not on unobserved measurements.
- MNAR (missing not at random): Missingness depends on unobserved measurements conditional on observed measurements.

Joint modelling is closely linked with dropouts in that dropout time can be considered as survival outcome. A clear distinction is that, from missing data perspective, the dropout is typically inferred from an individual's failing to be present at a scheduled follow-up time due to some reasons and considered as discrete time outcome, whilst time-to-event of interest is recorded or censored at the study end-time from joint modelling perspective. Thus, there is an association between the dropout and the underlying longitudinal trajectories and this can be considered missing not a random (MNAR), which can potentially affect the conclusions in case of ignoring the missing data in longitudinal analysis.

1.4 Challenges

As with any emerging research area, joint modelling has some statistical and methodological challenges which need to be addressed. Rizopoulos (2012)

claims that sample size and power calculations, as well as the general design of joint modelling have still a long way to go to unify the research methodology to the same levels seen in classic regression models.

The standard closed-form sample size calculations cannot be applied to the joint models. Numerical simulation is the only viable tool due to the complexity of joint models. Multiple visit times of subjects, as well as different number of visiting times and causes of dropouts need to be considered carefully. This research starts with considering the standard joint modelling assumptions of a single event of interest and a linear mixed effect model for longitudinal measurements, and extend the methodology by relaxing the assumptions. Considering more complex random effect structure leads to further technical issues in joint modelling. The approach of this study is a hybrid of theory and simulation.

It is worth noting that several important situations have yet to receive the requisite attention in joint modelling literature. For example, the treatment of asynchronous longitudinal profiles, multi-level data structures, sample size and power calculations, residuals and model fitting, dynamic prediction, model selection, handling of large multivariate datasets are all yet to be addressed. We will address some of these issues in this dissertation. Some of these problems can be dealt with simulation studies and comparisons with existing research studies.

1.5 Case studies

In this section, three datasets that contain different characteristics that the rest of the thesis seeks to explore are introduced, namely a multicentre dataset (albeit with a small sample), a multivariate longitudinal dataset with over thirty biomarkers, and a dataset with atypically long follow-up.

1.5.1 SLS data

The Scleroderma Lung Study, a 13-centre double-blind, randomized, placebo-controlled trial sponsored by the National Institutes of Health, was designed to evaluate the effectiveness and safety of one-year oral intake of cyclophosphamide (CYC) in patients with active, symptomatic scleroderma-related interstitial lung disease (Tashkin et al., 2006). Data was available from 158 patients, equally distributed into two treatment groups, namely CYC and placebo, and followed-up for two years. The study included four different types of events. There were 4 deaths, 12 treatment failures, 18 adverse events (AE) or serious adverse events (SAE), and 16 participants were withdrawn from the study without AE or SAE. The rest of the patients experienced no events. The event types and their frequencies are indicated in Table 1.1. The primary outcomes of this study are repeated measurements of the FVC (forced vital capacity, % predicted) and death/failure times. The dataset was used as an application in the textbook of joint modelling written by Elashoff et al. (2016), and is publicly available after registering at `https://faculty.biostat.ucla.edu/gangli/jm-book-data`, and was downloaded from the aforementioned website on December 15, 2016.

Having multiple centres makes this dataset a good application while investigating a multilevel joint model in Chapter 3. The dataset has tied event times, and such ties were broken by subtracting a tiny random value from each tied survival time as suggested by Borucka (2014). Ties in survival data is explained further in Section 2.3.3. In this work, some criteria were set for this dataset. Throughout this thesis, event values [1], [3] and [4] are pooled as having the event, and event values [0] and [2] as having no event, to create a new event indicator, yielding 34 events. Furthermore, 21 measurements taken after having the event from the seven patients were discarded. Five patients who had missing maximum fibrosis were also discarded. Lastly, 17 patients who did not complete the treatment in the first six months were discarded. In total, 22 patients who did not comply with the criteria (eight of whom had the event)

Type of event [Values]	Frequency
No event [0]	108
Adverse events (AE) and serious adverse events (SAE) [1]	18
Withdrawn due to non-response (dropouts without AE and SAE) [2]	16
Treatment failure [3]	12
Death [4]	4

Table 1.1: Frequencies of the event types for the SLS data

were discarded. Thus, 136 patients out of 158 were analysed, 26 of whom had the event, and 110 of whom did not have event in the final dataset. Hence, the event rate of this study is 19.1%. Figure 1.1 shows a plot of the profile

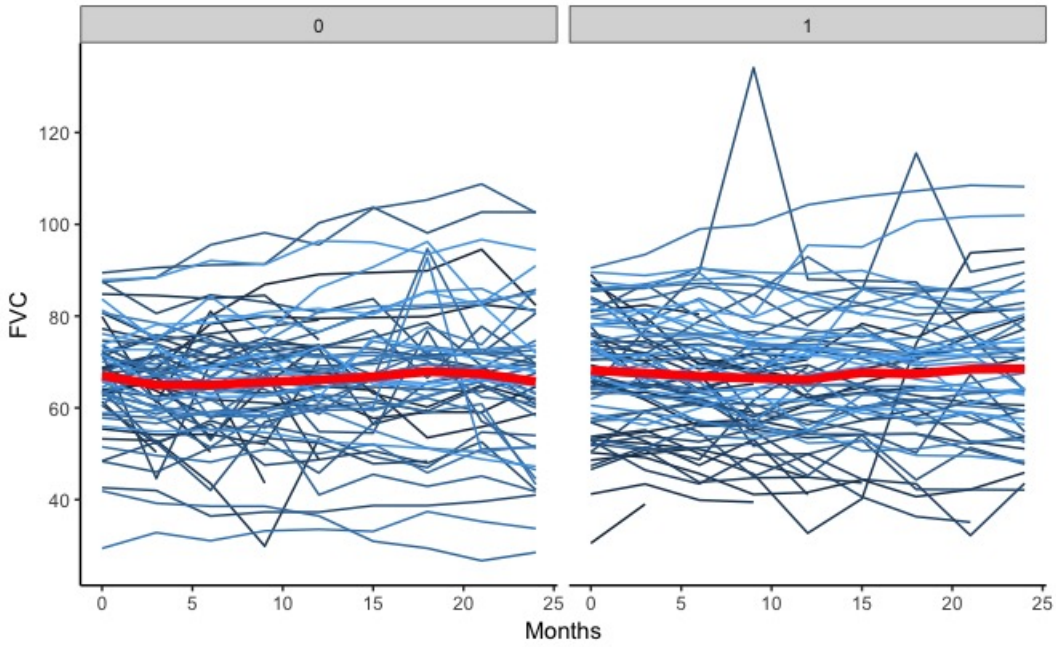

Figure 1.1: Profile plots of FVC% for CYC (0) group vs. placebo (1) group. Red line shows (smoothed) mean index

of FVC% stratified by treatment group. There is a large variation in baseline FVC% suggesting individual heterogeneity. Most patients showed a small variation in FVC% over time, however, there were a few patients with larger or smaller measurements. The bold red line indicates the mean profile of FVC%. It seems that the mean profiles of the two treatment types show no real differ-

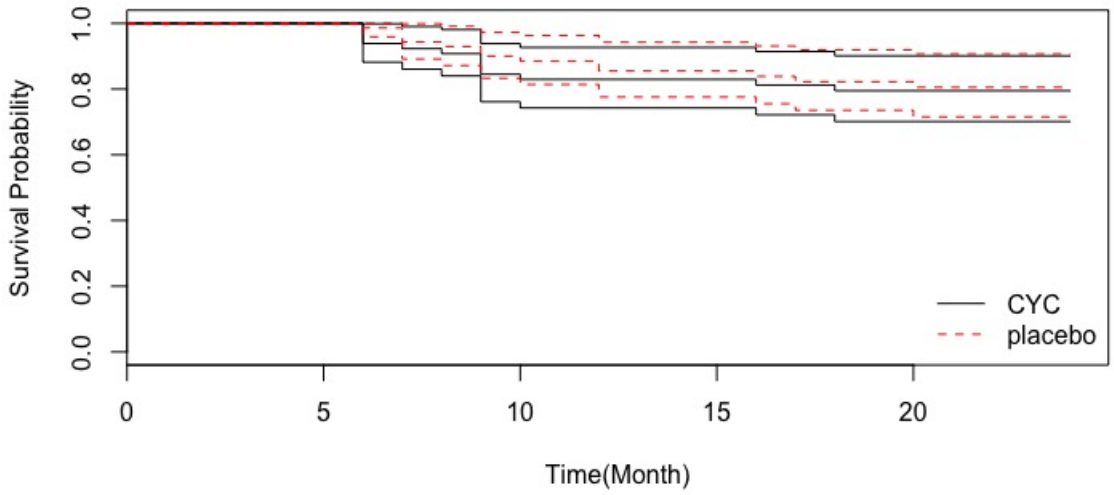

Figure 1.2: Patient survival with 95% CI bands on the SLS study

ence. Figure 1.2 shows the Kaplan-Meier survival curves. As can be seen, there is little difference between treatment groups in survival during the follow up time, and the placebo arm seems to be a slightly better treatment choice.

1.5.2 Liver cirrhosis trial

A study investigating the effect of prednisone treatment in patients with liver cirrhosis is another motivating dataset used throughout this thesis. This study was originally described by Andersen et al. (1993), and an application using prothrombin index as a surrogate biomarker for overall health was made by Henderson et al. (2002). Data is available for 488 patients with 40% censoring rate. Furthermore, 251 patients were randomly allocated at diagnosis to prednisone treatment, and 237 patients were randomised to placebo only. The patients were followed until death or the end of the study. In total, 292 patients died during the study. The study lasted 13 years, which is atypically long for a longitudinal study. The primary interests are the prothrombin index, measured repeatedly throughout the study and treatment. This dataset is a standard joint modelling dataset, with exception of the long follow-up time. In the first 2-3 years, there was little difference between treatment groups in survival.

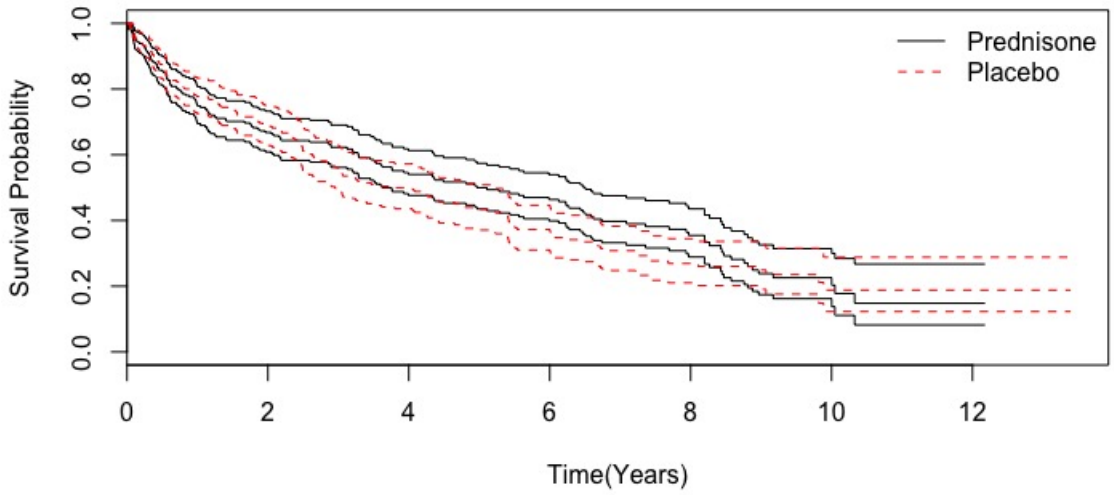

Figure 1.3: Survival probability of liver cirrhosis trial per treatment group with associated 95% CI bands.

The curves then diverge, which indicates a possible improved prognosis in the treatment arm (see Figure 1.3). The curves subsequently come together again around nine years into the study.

Figure 1.4 shows some exploration of the observed data, by plotting the longitudinal profile of the prothrombin index for each patient who took the placebo (left panel) or prednisone (right panel) treatment. As can be seen, the plot does not show much difference between the two treatments. The data is publicly available from the `JM` package in `R` (Rizopoulos, 2010).

1.5.3 ADNI data

The Alzheimer's Disease Neuroimaging Initiative (ADNI) study is a longitudinal multisite observational study of elderly individuals with normal cognition, mild cognitive impairment (MCI), or Alzheimer's disease (AD) (Mueller et al., 2005b,a), jointly funded by the National Institute of Health (NIH). The study is designed to assess the performance of magnetic resonance imaging (MRI), (18F)-fludeoyglucose positron emission tomography (FDG PET), urine, serum,

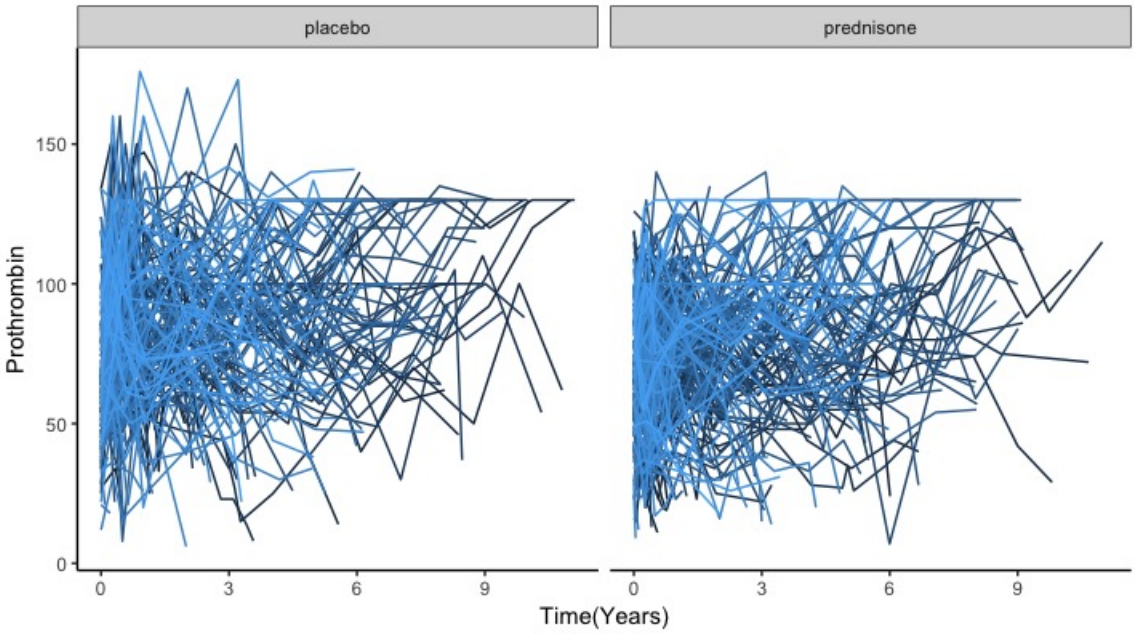

Figure 1.4: Longitudinal profiles of prothrombin index per treatment group for patients who were censored/died.

and cerebrospinal fluid (CSF) biomarkers, as well as various clinical and neurocognitive measures in terms of investigation into the progression of AD in the three groups of elderly individuals (Jack et al., 2008). More detailed information regarding the study's procedures and inclusion and exclusion criteria is available at `http://www.adni-info.org`. The data used in this thesis is publicly available after registering on the aforementioned website, and was downloaded from `http://ida.loni.ucla.edu` on October 24, 2017.

The dataset presented here includes 388 patients with MCI at baseline having at least one follow-up visit, and having the protocol from which subject originated is "ADNI-I". The criteria for MCI defined by Petersen et al. (1999) and used by Li et al. (2017) are the same as those adopted in this thesis: a memory complaint where objective memory loss is measured by education adjusted scores on the Wechsler Memory Scale Logical Memory II, a Folstein Mini Mental State Examination Score (MMSE) of 24-30, a Clinical Dementia Rating (CDR) equal to 0.5, absence of significant levels of impairment in functional, behavioural and neuroimaging domains, and essentially preserved activities of their routine. All subjects were given a consent form at the beginning of the

study, and the study had ethical approval from the local institutional review board at all sites that participated in the study. Having multiple longitudinal biomarkers makes this dataset a unique application for this thesis along with the investigation of multivariate joint modelling in Chapter 6.

The event of the study considered throughout this thesis is conversion to AD from MCI. There were 205 events recorded for the 388 patients, resulting in an event rate of 52.8%. This study has 33 longitudinal clinical and imaging measures, and is a rich, multivariate dataset. Figure 1.5 depicts the Kaplan-Meier survival curves of the ADNI-I study with the confidence intervals. Information about the baseline characteristics of the ADNI-I subjects with MCI are given in Chapter 6.

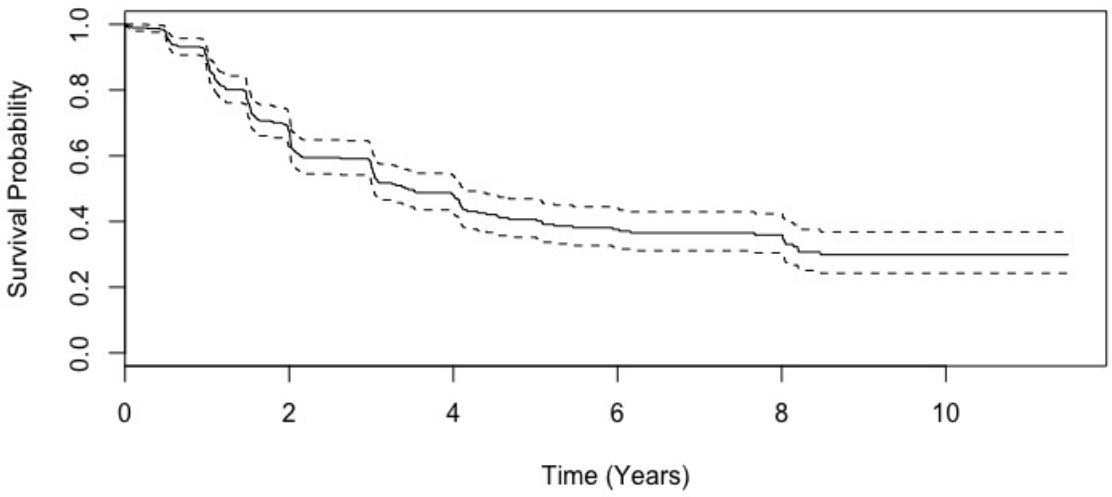

Figure 1.5: Survival probability of ADNI data with associated 95% CI bands.

1.6 Aim and objectives

The aim of this research is to extend the current statistical methodology design and incorporate multi-level models and multivariate joint models in order to improve the predictive ability of inference in joint modelling frameworks.

This aim begets the following objectives:

1. To extend the joint modelling framework to allow for multiple levels (such as the hospital or clinic where a patient was treated or a family effect), and to explore the consequences of ignoring the centre level effect when it is actually needed. An application of this is to use the SLS data discussed in the previous section, with rigorous simulation studies used to demonstrate the properties of the multilevel model.

2. To assess the power of the score test for association between biomarker values and survival time, both in the context of the classic Wulfsohn and Tsiatis (1997) joint model and extensions relevant to this work, namely joint models with multilevel latent associations, where in each case, association is via a single parameter (i.e. the univariate score test). Simulation studies are utilised to verify findings.

3. To develop an assessment of the power of a multivariate score test for association between longitudinal biomarkers and survival time when separate associations are considered for each random effect, to test whether a fit of joint models with multilevel random effect structure is needed, and to assess which biomarker is required to fit in the case of having multiple biomarkers. Simulation studies are performed for a range of scenarios to indicate how powerful the score test is under these situations.

4. To demonstrate how joint modelling can be used to maximise the information obtained from multiple, correlated longitudinal measurements, in order to improve the predictive ability of multivariate joint models over the univariate joint model and what gain can be made in moving from a univariate joint model to multivariate joint model and investigate the dynamic nature of the longitudinal profiles for a new patient. A particular application involves ADNI data, as introduced in the previous section.

5. To assess the sensitivity and robustness of the joint models with different sample sizes. Simulation studies are conducted to demonstrate the effectiveness of the model.

6. To assess the performance of univariate joint models of longitudinal and survival data over a separate analysis of these two components, and to investigate the consequences of fitting misspecified joint models to data that is generated under three different latent association structures. Simulation studies with various scenarios and latent association structures are drawn upon to verify findings, and fitting joint models are utilised for the liver cirrhosis data, introduced earlier.

7. To critically appraise all findings, suggest avenues for future research, and conclude on the achievements and challenges of the research appropriately.

1.7 Layout of the thesis

The thesis is organised as follows. Chapter 2 introduces the joint modelling of longitudinal and time-to-event data, assesses the performance of this technique when compared to the separate analysis of each component, and examines the impact on misspecified random effects structure in this framework. Chapter 3 develops multilevel joint models that account for centre level effect as well as individual effects, and investigates the consequences of fitting a joint model with individual level random effects when the centre level effect is ignored. Chapter 4 introduces the score test for association between the longitudinal and survival data, and performs simulation studies under various scenarios. Chapter 5 extends the score test for association between these two components with various situations, i.e. univariate joint model in conjunction with a separate association, multilevel joint model, and multivariate joint model. Chapter 6 introduces the joint modelling of multivariate longitudinal and survival data as well as prospective accuracy and dynamic features of londitudinal biomarkers, while Chapters 7 presents an assessment of the robustness of the univariate joint model. Finally, Chapter 8 concludes the thesis and provides recommendations for future work.

Chapter 2

Joint modelling of longitudinal and time-to-event data

2.1 Introduction

Although repeated measurements and survival time data are often collected in tandem (e.g. the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog), and time until the conversion of MCI to AD for the ADNI study explained in Section 6.3 in detail (Li et al., 2017)), they are often analysed separately due to the lack of availability of suitable software before the emergence of the `JM` package in `R` (Rizopoulos, 2010), and lack of penetration of joint modelling into other disciplines. A more complex approach involves a requirement to analyse these two kinds of associated measurements in a more efficient way. This leads to the joint modelling of longitudinal and time-to-event data, which has seen an explosion of interest in the past twenty years. The aim of joint modelling is to investigate and exploit any potential association between longitudinal measurements and time-to-event data. Joint modelling reduces bias and makes the analysis more efficient by using the available data in an optimum way (Asar et al., 2015). A common approach involves a combination of linear mixed effect models and Cox proportional hazards model for the longitudinal and survival submodels, which can then be linked through the shared random effects. Tsiatis and Davidian (2004) illustrated the association between longitudinal and time-to-event data, and gave one of the most common examples used in joint

modelling literature. In this example, HIV studies, CD4 count cells are measured repeatedly over time as measures immunologic and virologic status, and patients are also followed-up until they contract AIDS or die.

The primary focus for inference on joint modelling may depend on the application of interest. There are typically three objectives for joint modelling:

1. Inference about the repeated measurements with possible non-ignorable dropout; or
2. The distribution of time-to-event of interest conditional on time-varying longitudinal outcome; or
3. The joint relationship between the longitudinal outcomes and time-to-event process.

For example, Henderson et al. (2000) analysed the PANSS score with the consideration to be important, with equal interest in both longitudinal and event time data, third objective.

In joint modelling, full likelihood methods to estimate the parameters from the joint distribution of the two outcomes can differ depending on the factorisation of the joint distribution of the components (Sousa, 2011; McCrink et al., 2013). Bayes' rule is applied for the factorisation of the joint distribution. The two different factorisations of the joint distribution require different modelling strategies in terms of inference of the model. They are referred to as pattern-mixture models and selection models (Little and Rubin, 2014; Little, 1993). If Y denotes longitudinal outcomes, and T denotes the survival outcomes, the joint distribution of these two outcomes can be formulated as:

$$\text{Pattern-mixture models:} \quad [Y,T] = [T][Y|T] \tag{2.1}$$

$$\text{Selection models:} \quad [Y,T] = [Y][T|Y] \tag{2.2}$$

Although the joint distribution of these two models are the same, they have different interpretations. The nature of the statistical problem can affect the choice of the models. Pattern-mixture models deal with the problem when the

primary interest is the longitudinal outcomes of different patterns of drop-out, which can also have the association between event times. If the inference is on the event times parameters, selection models are used to estimate the model parameters. Simple, yet detailed explanations on these models are provided by Sousa (2011) and McCrink et al. (2013). Furthermore, Little (2008) and Elashoff et al. (2016) explain these models in Chapter 4 in detail.

The natural extension of these models is the incorporation of random effects, which makes the previous models known as random pattern-mixture models and random selection models. The marginal longitudinal models include individual random effects, which are unobserved in real problems in the former model, while the marginal distribution of event times are included in the latter model. The models are expressed as: (Sousa, 2011):

$$\text{Random pattern-mixture models:} \quad [Y, T, U] = [U][T|U][Y|T] \tag{2.3}$$

$$\text{Random selection models:} \quad [Y, T, U] = [U][Y|U][T|Y] \tag{2.4}$$

where U denotes the random effects.

Another different class of joint models is defined by Diggle (1998) as random effects models. The assumption of this model is that both longitudinal outcomes and event times are dependent on the individual unobserved random effects, with a further assumption of conditional independence between repeated measurements and event times, given the random effects. The model is defined as:

$$\text{Random effects models:} \quad [Y, T, U] = [U][Y|U_1][T|U_2] \tag{2.5}$$

where $U = (U_1, U_2)$. In these models, the two processes are independent, conditional on the random effects, and the association between these outcomes is determined by the correlation structure between U_1 and U_2.

The third objective is the primary focus for inference, as the primary interest of this thesis is random effects models. Hence, a detailed explanation of the statistical methods is first provided for the linear mixed-effects model, survival analysis and joint models. Simulation studies are then conducted, followed by

an application of the liver cirrhosis data. Lastly, this chapter is concluded with a brief discussion.

2.2 Linear mixed-effects model

In longitudinal studies, serial observations on the same units are taken. A suitable analysis for such data needs to take this potential dependence within individuals into account. The linear mixed-effects model (LMEM) is a very effective and widely used method to deal with unbalanced longitudinal data structures, which have different number of observations for each subject taken at potentially different times (Laird and Ware, 1982). The LMEM includes three parts, namely fixed effects, random effects and random errors. The underlying idea of the LMEM is that regression coefficients may vary randomly from one subject to another. For instance, different subjects may have different intercepts and slopes.

Notationally, the LMEM can be outlined as follows. Let Y_{ij} denote the j^{th} measurements of the i^{th} subject taken at a pre-specified time point of t_{ij}. There are m_i measurements taken for the i^{th} subject. That is, $j = 1, 2, \ldots, m_i$ and $i = 1, 2, \ldots, n$, where n is the number of subjects. The model formulation can then be expressed as follows:

$$Y_{ij} = \boldsymbol{x_{1i}}(\boldsymbol{t'_{ij}})\boldsymbol{\beta_1} + W_{1i}(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n \tag{2.6}$$

$$W_{1i}(t_{ij}) = D(t_{ij})' U_i, \quad U_i \sim N(\mathbf{0}, \Sigma), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where $W_{1i}(t_{ij})$ is a latent process incorporating random effects, U_is and ε_{ij}s are assumed to be mutually independent, $\boldsymbol{\beta_1}$ is a p-vector of fixed-effects coefficients, $\boldsymbol{x_{1i}}(\boldsymbol{t_{ij}})$ is m_i-vector of covariates for individual i, Σ is a $q \times q$ positive-definite variance-covariance matrix, and $D(t_{ij})$ is a vector of explanatory variables for individual i. To borrow from Crowder (2017), "To be poetic, $\boldsymbol{x_{1i}}(\boldsymbol{t'_{ij}})$ is an immutable constant of the Universe, $W_{1i}(t_{ij})$ is a lasting characteristic of the individual, and ε_{ij} is but a fleeting aberration of the moment".

The advantages of mixed models include the possibility to estimate parameters describing the response changes in the population of interest as well as the ability to predict the changes in the subject response trajectories over time. The overall variability is partitioned between subjects and random errors. Furthermore, inference at the subject level is one of the main reasons for using mixed models in the joint modelling of longitudinal data analysis and time-to-event analysis.

2.3 Survival analysis

Survival data are of particular interest in a variety of applied fields such as medicine and biology, where occurrence of certain events times or time to death from the onset of a disease are of particular interest. There are four fundamental functions in a time-to-event analysis, namely the cumulative distribution function, $F(t)$, the survival function, $S(t)$, the hazard function, $h(t)$, and the cumulative hazard function, $H(t)$. Mathematically, they are intrinsically related to each other. Let T denote the time to occurrence of an event of interest, and suppose T has a probability distribution with underlying probability density function, $f(t)$. One can then express the following:

- The cumulative distribution function:

$$F(t) = P(T \leq t) = \int_0^t f(u)du \tag{2.7}$$

- The survival function:

$$S(t) = P(T > t) = 1 - F(t), \quad t \geq 0 \tag{2.8}$$
$$\frac{dS(t)}{dt} = -f(t)$$

- The hazard function:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P[(t \leq T < t + \Delta t) | T \geq t]}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{P(T < t + \Delta t) - P(T \leq t)}{S(t)} \\ &= \lim_{\Delta t \to 0} \frac{1}{S(t)} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{F'(t)}{S(t)} \\ h(t) &= \frac{f(t)}{S(t)} \end{aligned} \tag{2.9}$$

- The cumulative hazard function:

$$\begin{aligned} H(t) &= \int_0^t h(u) du \\ &= \int_0^t \frac{f(u)}{S(u)} du = \int_0^t \frac{F'(u)}{S(u)} du = -\int_0^t \frac{S'(u)}{S(u)} du = -\ln[S(t)] \\ H(t) &= -\ln S(t) \\ S(t) &= \exp(-H(t)) \end{aligned} \tag{2.10}$$

The survival and the hazard functions are the most useful functions in terms of explaining risk. The survival function gives the probability of experiencing the event of interest within a specified time frame, while the hazard function gives the rate of experiencing the event per year.

2.3.1 Parametric survival analysis

In survival analysis, using a parametric model can help to gain a greater insight into the observed data, as a result of its flexibility and the fact that it is better for standard errors of parameters. These models, while not considered explicitly here, will be subsequently used to illustrate concepts. They are usually based on the hazard or the log-hazard function, where the hazard can either increase or decrease with time. The assumptions that was made about the shape of the hazard function can specify the time-to-event. There are some parametric distributions in the hazard functions such as exponential, Weibull or Gompertz. The hazard, survival and density functions of these distributions are given in the following table, where $t > 0$.

Functions	$h(t)$	$S(t)$	$f(t)$
Exponential Distribution	λ	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$
Weibull Distribution	$\lambda\gamma t^{\gamma-1}$	$\exp(-\lambda t^{\gamma})$	$\lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma})$
Gompertz Distribution	$\lambda \exp(\gamma t)$	$\exp\{-\lambda\gamma^{-1}(e^{\gamma t}-1)\}$	$\lambda \exp\{\gamma t - \lambda\gamma^{-1}(e^{\gamma t}-1)\}$

Table 2.1: Some common lifetime distributions

The simplest assumption that one can make about the shape of the hazard function is having a constant hazard rate over time, which assumes that the survival times follow an exponential distribution. The Weibull distribution is a more flexible choice for the hazard function, where the hazard rate increases or decreases monotonically. Another parametric distribution is the Gompertz distribution, which is more likely to be used in mortality data. In such a distribution, the hazard rate increases or decreases exponentially.

2.3.2 Models for survival analysis

To motivate ideas, the estimation of the survival function of the Kaplan-Meier estimator will be explained, as well as the cumulative hazard function of the Nelson-Aalen estimator. The Cox proportional hazards model will then be discussed.

Kaplan-Meier (KM) estimator

Proposed by Kaplan and Meier (1958), the KM estimator is the most well-known method for comparing survival functions for a small number of groups. It is a non-parametric method that does not require any assumption underlying the distribution of the failure times. To introduce this estimator, let $t_1 < t_2 < \cdots < t_k$ denote the observed event times in a sample of n subjects, and note that some subjects may be censored. The survival function of the KM estimator is defined by:

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{r_i}\right) \tag{2.11}$$

where r_i is the number of subjects at-risk at time t_i, d_i is the number of events at time t_i. The variance of the above survival function can be calculated using Greenwood's formula.

$$Var(\hat{S}_{KM}(t)) = \hat{S}_{KM}(t)^2 \sum_{t_i \leq t} \left(\frac{d_i}{r_i(d_i - r_i)} \right) \tag{2.12}$$

Nelson-Aalen (NA) estimator

Originally suggested by Nelson (1972) and studied by Aalen (1978), the NA estimator is used to estimate the cumulative number of expected events in a given time. It is an alternative approach to the KM estimator and can be thought of as a non-parametric method to estimate the cumulative hazard function. The NA cumulative hazard rate estimator is given by:

$$\hat{H}_{NA}(t) = \sum_{t_i \leq t} \frac{d_i}{r_i} \tag{2.13}$$

the survival function based on the NA estimator, using the relation (2.10), can be calculated as:

$$\hat{S}_{NA}(t) = \exp\left\{ -\hat{H}_{NA}(t) \right\} = \prod_{t_i \leq t} \exp\left(\frac{d_i}{r_i} \right) \tag{2.14}$$

These two estimators of the survival function are asymptotically equivalent. The performance of the NA estimator is slightly superior to that of the KM estimator. However, the KM estimator performs better in terms of decreasing failure rates, while the NA estimator provides better results for increasing failure rates (Colosimo et al., 2002). An illustration of the KM estimator of the survival function and the NA estimator of the cumulative hazard function for the liver cirrhosis data is depicted in Figure 2.1. The survival curves in the left panel of Figure 2.1 show that less than 20% of patients survive 10 years after entry. The prednisone-treated patients have slightly better prognosis than those who were given placebo between around 2 and 10 years into the study.

2.3.3 The Cox proportional hazards model

Time-to-event analysis investigates the relationship of the time-to-event distribution to one or more, possibly time-varying, covariates. Usually, specification

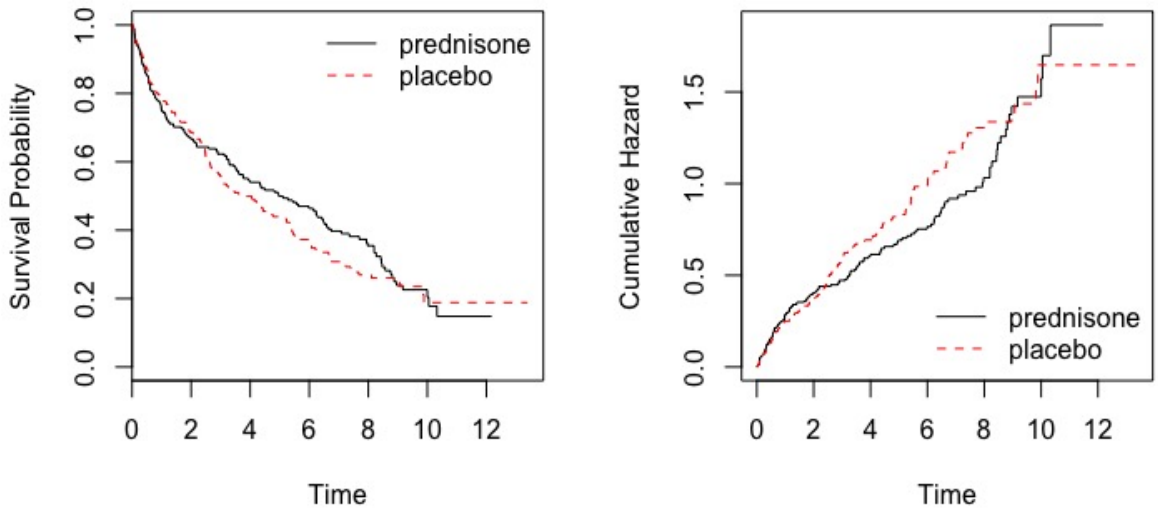

Figure 2.1: Left panel - Kaplan-Meier estimate of the survival function for the liver cirrhosis data. Right panel - Nelson-Aalen estimate of the cumulative hazard function for the liver cirrhosis data.

of a linear-like model is adopted for the log-hazard. For example, a parametric model based on the exponential distribution can be expressed as:

$$\log h_i(t) = \alpha + \boldsymbol{x_{2i}}\boldsymbol{\beta_2} \tag{2.15}$$

or equivalently,

$$h_i(t) = \exp\{\alpha + \boldsymbol{x_{2i}}\boldsymbol{\beta_2}\} \tag{2.16}$$

where $h_i(t)$ shows the hazard rate for the ith subject at time t, $\boldsymbol{x_{2i}}$ is the covariates vector (they do not change over time in this case), $\boldsymbol{\beta_2}$ are the associated regression coefficients that need to be estimated, and α represents a kind of log-baseline hazard, since $\log h_i(t) = \alpha$ when all $\boldsymbol{x_{2i}}$ elements are zero. It can be said that equation (2.15) is a linear model for the log-hazard or equation (2.16) is a multiplicative model for the hazard.

In contrast, in the Cox model, the baseline hazard function $\alpha(t) = \log h_0(t)$ is unspecified. Equation (2.15) is then written as:

$$\log h_i(t) = \alpha(t) + \boldsymbol{x_{2i}}\boldsymbol{\beta_2} \tag{2.17}$$

or equivalently,

$$\begin{aligned} h_i(t) =& h_0(t)\exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2}\} \\ =& h_0(t)\exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2}) \end{aligned} \tag{2.18}$$

where $h_0(t) = \exp\{\alpha(t)\}$.

The Cox model is a semi-parametric model. Whilst it does not require any assumption about the shape of the baseline hazard, $h_0(t)$, it requires some restrictive assumptions, one of which concerns tied events (i.e. events with exactly the same time). While $h_0(t)$ can take any form of baseline hazard, the covariates enter the model linearly. The key assumption of equation (2.18) is that covariate effects remain constant over the follow-up time. Cases with tied event times never happen if time was measured in a perfectly continuous scale. However, in practice, time is generally measured in a discrete manner, which is more likely to exist in tied event times. Throughout this thesis (as in the SLS data), the ties in survival data are broken by subtracting a small random value from each tied survival time to prevent the violation of this assumption (Borucka, 2014). This simple method provides results that do not differ from the exact results to a great extent.

Now, consider two individuals, i and i^*. The hazard ratio is:

$$\frac{h_i(t)}{h_{i^*}(t)} = \frac{h_0(t)\exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2})}{h_0(t)\exp(\boldsymbol{x_{2i^*}}\boldsymbol{\beta_2})} = \exp(\boldsymbol{x_{2i}} - \boldsymbol{x_{2i^*}})\boldsymbol{\beta_2} \tag{2.19}$$

which is independent of time t. Since (2.19) is independent of the hazard, the Cox model is a proportional hazards (PH) model.

Cox (1972) derived a partial likelihood function for the i^{th} subject for a PH model, which can be expressed as follows:

$$L_i(\boldsymbol{\beta_2}) = \left[\frac{\exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2})}{\sum_{i\epsilon R(t_i)}\exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2})}\right]^{\delta_i} \tag{2.20}$$

where t_i and δ_i are the event/censoring time and the indicator for the i^{th} individual, respectively. $R(t_i)$ is the set of individuals who are at risk at time t_i, that

is $R(t_i) = \{j : t_i \leq t_j\}$. Crucially, this does not depend on the baseline hazard, only the order of the event times.

The log-likelihood function can be written as:

$$\log L_i(\boldsymbol{\beta_2}) = \delta_i \Big[\boldsymbol{x_{2i}} \boldsymbol{\beta_2} - \log \sum_{i \epsilon R(t_i)} \exp(\boldsymbol{x_{2i}} \boldsymbol{\beta_2}) \Big] \tag{2.21}$$

Equation (2.20) is not a full likelihood, rather, it is called a partial likelihood (Cox, 1975), since it does not use any observed survival or censoring times. The vector of parameters $\boldsymbol{\beta_2}$ can be estimated by the Newton-Raphson method.

In the PH model, it is assumed that the hazard depends only on the constant covariates during the follow-up time, such as age, sex and treatment indicator. Time-dependent covariates can be of interest in many studies, especially regarding the association with the risk of an event, such as environmental factors, biochemical parameters or adjustment to treatment dose. Crowther (2014) developed extensions to the Cox model to allow for time-dependent covariates.

2.4 The standard joint model

In the standard joint model, a LMEM for the repeated measurements and the Cox PH model for the time-to-event are considered. To introduce the standard joint model, let T_i^* denote the *true* event time for the i^{th} subject, T_i denote the observed event time and $T_i = min(T_i^*, C_i)$, where C_i is the censoring time, and let $\delta_i = I(T_i^* \leq C_i)$ denote the event indicator (if the survival time is censored for the i^{th} subject $\delta_i = 0$, otherwise $\delta_i = 1$). Furthermore, let $Y_i(t)$ denote the observed endogenous time-dependent covariate at time point t for the i^{th} individual. However, it should be noted that $Y_i(t)$s are not actually observed, instead at a particular occasions t_{ij} the observed longitudinal measurements $Y_{ij} = \{Y_i(t_{ij}), j = 1, \cdots, m_i\}$ are obtained. The standard formulation of a joint model consists of two separate submodels, namely the longitudinal submodel and the survival submodel.

Define $W(t) = \{W_1(t), W_2(t)\}$ as a latent bivariate Gaussian process and assume that the time varying risk factors, namely the longitudinal measurements and time-to-event, are conditionally independent, given $W(t)$ and the other covariates. Therefore, the association is described through the latent processes, $W_1(t)$ and $W_2(t)$.

2.4.1 The longitudinal submodel

The main idea of a joint model is to measure the association between the longitudinal biomarker and the risk of an event, through some shared parameters. W_{1i} is introduced to denote a latent process which incorporates random effects, and possibly, a stationary Gaussian process. The longitudinal measurements are taken intermittently with error. In order to estimate the true and unobserved longitudinal outcomes, for each subject, the construction of a LMEM to describe the subject-specific trajectory function is required. The LMEM is:

$$\begin{aligned} Y_{ij} &= \boldsymbol{x_{1i}}(\boldsymbol{t_{ij}})'\boldsymbol{\beta_1} + W_{1i}(t_{ij}) + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \sigma_\varepsilon^2) \end{aligned} \tag{2.22}$$

where $\boldsymbol{x_{1i}}(\boldsymbol{t_{ij}})'$ is a set of baseline covariates for the fixed-effects, $\boldsymbol{\beta_1}$ is the vector of corresponding regression coefficients, W_{1i} is a latent process, and ε_{ij} is a sequence of mutually independent measurement errors.

2.4.2 The survival submodel

The proportional hazard submodel is given below in order to understand the strength of the association between the longitudinal outcome and the risk for an event.

$$\begin{aligned} h_i(t|\boldsymbol{x_{2i}}, W_{2i}) &= \lim_{dt\to 0} Pr\{t \le T_i^* < t + dt | T_i^* \ge t, \boldsymbol{x_{2i}}, W_{2i}\}/dt \\ &= h_0(t)\exp\{\boldsymbol{x_{2i}}(\boldsymbol{t})\boldsymbol{\beta_2} + W_{2i}(t)\}, \quad t > 0, \\ W_{2i}(t) &= \gamma W_{1i}(t) \end{aligned} \tag{2.23}$$

where $h_0(\cdot)$ is the baseline hazard function, $\boldsymbol{x_{2i}}(\boldsymbol{t})$ is a vector of a set of baseline covariates (such as sex, age and treatment indicator), $\boldsymbol{\beta_2}$ is the vector of corresponding regression coefficients, and γ is the association parameter. $\boldsymbol{x_{2i}}$ may

or may not have a common element with $\boldsymbol{x_{1i}}$. Involving $W_{2i}(t)$ into the linear predictor of the proportional hazard model links two submodels in the joint modelling framework. Equation (2.23) shows that the risk depends only on the current value of longitudinal outcome:

$$\begin{aligned} S_i(t|\boldsymbol{x_{2i}}, W_{2i}) &= Pr(T_i^* > t|\boldsymbol{x_{2i}}, W_{2i}) \\ &= \exp\Big(- \int_0^t h_0(s) \exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + W_{2i}(s)\} ds \Big), \end{aligned} \tag{2.24}$$

From equation (2.24), the survival function is based on the whole history of longitudinal trajectory. The time-dependent feature of the survival function requires numerical integration, which will be explained further.

In the standard survival analysis, it is widely popular to leave the baseline hazard function completely unspecified in order to avoid making any assumptions about the distribution of the survival times (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000). Hsieh et al. (2006) states that in the joint modelling framework, an unspecified baseline hazard function may lead to underestimated standard errors of the corresponding parameters. This problem can be avoided with the definition of $h_0(\cdot)$, or taking a parametric distribution instead of semi-parametric Cox model, such as the exponential, the Weibull, the Gompertz, or the log-normal distribution. The choice of a parameter-flexible specification of the baseline hazard function can be another option to avoid the underestimation problem.

2.5 Choice of latent process

Henderson et al. (2000) referred to the direct link between $W_1(t)$ and $W_2(t)$ as a latent association. In the case of having no latent association, there is no benefit of the joint analysis. There is a wide range of specific models with the combination of Equations (2.22) and (2.23). In particular, Henderson et al. (2000) proposed a flexible choice for $W_{1i}(t)$ by combining the suggestions in Laird and Ware (1982) and Diggle (1988) as follow:

$$W_{1i}(t) = D_i(t)U_i + V_i(t) \tag{2.25}$$

where $D_i(t)$ is a vector of explanatory variables, $U_i \sim N(0, \sigma_1^2)$ is a vector of random effects, and $V_i(t)$ is a stationary Gaussian process with zero mean and variance σ_v^2, and with a correlation function of $r(u) = cov\{V_i(t), V_i(t-u)\}/\sigma_1^2$.

The random effects model mostly takes three different forms and they are referred as Models I, II and III in simulation studies for this chapter:

- Random intercept model (Model I): U_0
- Random intercept and slope model (Model II): $U_0 + U_1 t$
- Random quadratic model (Model III): $U_0 + U_1 t + U_2 t^2$

The latent association depends on the type of the random effects model. The most common four types of association are:

- Proportional association: $W_2(t) = \gamma W_1(t)$

$$h_i(t) = h_0(t) \exp\{\boldsymbol{x_{2i}(t)'\beta_2} + \gamma(U_{0i} + U_{1i}t)\}$$

- Subset association:

$$h_i(t) = h_0(t) \exp\{\boldsymbol{x_{2i}(t)'\beta_2} + \gamma U_{0i}\}$$

- Separate effect association:

$$h_i(t) = h_0(t) \exp\{\boldsymbol{x_{2i}(t)'\beta_2} + \gamma_0 U_{0i} + \gamma_1 U_{1i}t\}$$

- Frailty:

$$h_i(t) = h_0(t) \exp\{\boldsymbol{x_{2i}(t)'\beta_2} + \gamma_0 U_{0i} + \gamma_1 U_{1i}t + U_{2i}\}$$

where $U_{2i} \sim N(0, \sigma_2^2)$ is independent of (U_{0i}, U_{1i}), and U_2 models frailty orthogonal to the measurement process.

If $\gamma > 0$, positive values of the random effects increase the risk for the event of interest. Furthermore, the hazard increases as U increases. Similarly, if $\gamma < 0$, negative values of the random effects increase the risk for the event of interest.

2.6 Maximum likelihood estimation

Maximum likelihood estimation, proposed by Wulfsohn and Tsiatis (1997), is the primary estimation method used in the joint modelling literature. Here, the basics of the maximum likelihood method for the joint models are explained. The observed outcomes for i^{th} individual are $(T_i, \delta_i, \boldsymbol{Y_i}, \boldsymbol{t_i}, \boldsymbol{x_{1i}}, \boldsymbol{x_{2i}})$, representing the survival time, event indicator, longitudinal outcomes, measurement times, and baseline covariates in longitudinal and survival submodels, respectively. An assumption that can be made is that the vector of random effects U_i has been taken into account in both the longitudinal and event time processes, and the random effects are not observed. The observed data likelihood is;

$$\prod_{i=1}^{n}\left[\int_{-\infty}^{\infty}\left\{\prod_{j=1}^{m_i} f(Y_{ij}|\boldsymbol{U_i}, \sigma_\varepsilon^2)\right\} f(\boldsymbol{U_i}|\Sigma) f(T_i, \delta_i|\boldsymbol{U_i}, h_0, \boldsymbol{\gamma}, \boldsymbol{x_{2i}}) d\boldsymbol{U_i}\right] \tag{2.26}$$

where

$$f(Y_{ij}|\boldsymbol{U_i}, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-1/2}\exp\{-(Y_{ij} - \boldsymbol{x_{1i}}\boldsymbol{\beta_1} - \boldsymbol{D_i}\boldsymbol{U_i})^2/2\sigma_\varepsilon^2\}$$
$$f(\boldsymbol{U_i}|\Sigma) = (2\pi|\Sigma|)^{-1/2}\exp\{-(\boldsymbol{U_i})'\Sigma^{-1}(\boldsymbol{U_i})/2\}$$

and

$$f(T_i, \delta_i|\boldsymbol{U_i}, h_0, \boldsymbol{\gamma}, \boldsymbol{x_{2i}}) \tag{2.27}$$
$$= [h_0(T_i)\exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + \boldsymbol{\gamma}\boldsymbol{U_i}\}]^{\delta_i}\exp\left[-\int_0^{T_i} h_0(s)\exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + \boldsymbol{\gamma}\boldsymbol{D_i}\boldsymbol{U_i}\}ds\right]$$

Here, the common assumptions of independent censoring mechanisms and non-informative visiting process are made. The visiting process is the stochastic or deterministic mechanism generating the time points at which repeated measurements are taken.

The parametric maximum likelihood method estimates the parameters Σ, σ_ε^2, $\boldsymbol{\gamma}$, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$, while the nonparametric maximum likelihood method estimates $h_0(s)$. Let $\boldsymbol{\Omega}$ denote the whole set of these parameters. In order to estimate the parameters, the EM algorithm is used. The EM algorithm aims to estimate the parameters of interest by maximizing the likelihood function of the observed data. This procedure is iterated between E-step and M-step. E-step is

used when the expected log-likelihood of the complete data (the observed data and the random effects components for each individual) is conditional on the observed data, and computes the current estimates of the parameters, while M-step maximizes the expected log-likelihood to compute the new parameter estimates (Wulfsohn and Tsiatis, 1997). However, the linear convergence is a drawback of the EM algorithm. As an alternative, the estimation can be carried out via the Newton-Raphson algorithm, MCMC, or MCEM.

As noted in Rizopoulos et al. (2009), the score vector corresponding to the log-likelihood function can be calculated. The crucial feature of Equation (2.26) is that the integral does not have a tractable analytical solution, and as such, numerical techniques are needed to solve the integral, such as quadrature or Laplace approximation, or MCMC.

2.7 Simulation studies

Here, the performance of a selection of joint models is evaluated via simulation studies. Two simulation studies were conducted, the first of which compares the performance of the joint models over the separate analysis of each outcome. The second simulation study assesses the impact of fitting a misspecified joint model. Means and standard errors of the parameters of interest in simulation study I were calculated with the Monte Carlo method, whereas bootstrap method was used to estimate the parameters in simulation study II.

Two different techniques are used in order to provide an insight into using one technique instead of the other. Joint modelling itself is an intensive method and bootstrap requires resampling and it is also computationaly intensive. We would like to investigate if there is any gain using computationally intensive method to assess the performance of joint modelling in case of having limited computational access.

Monte Carlo simulation is generally considered a procedure that generates pos-

sible outcomes by sampling from a theoretical distribution with predefined parameters. It is simply a way to determine outcomes based on 1000s of theoretical return paths. Alternatively, bootstrapping, which can be thought of as a type of nonparametric Monte Carlo analysis, also runs thousands of simulations but takes the underlying data as given. Both of these methods can explore thousands of possible return paths, and derive confidence intervals encompassing these return paths. The major difference between the two is that Monte Carlo simulates data and bootstrapping takes the data as given and just resamples it over and over. What is advantageous about bootstrapping is that no assumption about the underlying distribution or its properties is assumed. However, bootstrapping does make the assumption that future paths will have the same basic historical return realizations that have been experienced in the past.

2.7.1 Simulation study I

In the first simulation study, data was generated, and joint models as well as separate models were used to fit the longitudinal and survival components. The primary objective of this simulation study is to assess how good the joint model estimates are, and how far the separate estimates are for the data linked with random effects. Two different scenarios were assumed, namely a 70% event rate and a 25% event rate, and are referred to here as Scenario I and Scenario II. Furthermore, three different random effects structures were adopted, namely random intercept (Model I), random intercept and slope (Model II), and random quadratic model (Model III) in conjunction with a separate association structure (i.e. a separate γ parameter for each random component). In special case of zero association between repeated measurements and time-to-event outcome, there is no benefit of fitting joint model. Joint model is degraded to separate analysis of each outcome.

In each of 500 repetitions, $n = 1000$ independent subjects were generated with discrete measurement times, having one continuous covariate drawn from the standard normal distribution, and one binary indicator drawn from $X_i \sim B(1, 0.5)$

for $i = 1, \cdots, n$. The repeated measurements were generated according to the longitudinal submodel, equation (2.22). Furthermore, the `simjoint()` function in `joineR` package was used to generate data (Philipson et al., 2017). This function simulates from a joint model (similar to the joint model in Henderson et al. (2000)) a range of latent association structures.

The event times were generated according to the survival submodel, equation (2.24). The same binary and continuous covariates as in the longitudinal submodel were also incorporated into the survival submodel. Moreover, censoring times were randomly generated with a positive scale parameter from an exponential distribution as default.

When the latent association of the model is a random intercept (Model I), the baseline hazard function is specified with an exponential distribution with

$$\lambda_0(t) = \exp\{\theta_0\}$$

whereas, when the latent association of the model is random intercept and slope (Model II), the baseline hazard model is specified to be a Gompertz distribution with

$$\lambda_0(t) = \exp\{\theta_0 + \theta_1 t\}$$

where θ_1 is the shape parameter, and $\exp\{\theta_0\}$ is the scale parameter. Here, the form of the hazard matches the latent structure, however, this is not a requirement. The methodology developed by Bender et al. (2005) was used for the simulation of the event times in Model I, whereas Austin (2012) was used for the simulation of the event times in Model II. The hazard can be a function of time in Model I, and can be independent of time in Model II, because the hazard is a separate part of the model. While these two models that match the latent structure were selected, the hazard does not have to follow the latent association structure, as it can be any type of hazard function, such as spline, piecewise, or linear.

When the latent association of the model is a random quadratic model, or if

the form of the hazard is nonlinear, Model III, the integral becomes intractable, making this approach different from that used in the above two models. By considering the hazard of failure at increasingly small time intervals, the instantaneous hazard rate can be written as:

$$h(t) = \lim_{dt \to 0} \lambda(t)dt = \lim_{dt \to 0} P[t \leq T \leq t + dt | T \geq t]$$

This new approach splits the time scale into nominally small increments, specified with the `gridstep` value, the default value of which is 0.01 in the `simjoint()` function. What this does is to act as dt in the integral, which multiplies the hazard. This hazard is equivalent to the probability of having event in the interval $(t, t + dt)$. For each individual, a vector of possible times is set from 0 to truncation time, possibly as the default maximum number of time points and the event probability, $h(t)$, is calculated and compared with a random probability drawn from $U(0, 1)$. If the probability is greater than the random draw, the failure time is set the time. If not, it is the generated time from the vector of candidate times. The event time is taken to be the minimum of the candidate times.

Suppose an individual is followed from time of enrolment, $t = 0$, in a study of a particular disease, until the subject dies from the disease. In addition, we assume that it is a 2 year study, measurements are taken monthly, and the individual is enrolled on the first day of the study. Furthermore, it is assumed that the time scale is split by 0.01 as default. The maximum length of follow-up for the individual is 24 months. First, a vector of length of $24 * 100 = 2400$ was generated, which is the maximum number of time points, denoted as t^*, $t^* = (24, 24, \cdots, 24)$. After comparing with a random probability, drawn from $u \sim U(0, 1)$, the vector of possible times becomes $t^* = (24, 24, 0.02, 24, 0.04, \cdots, 24)$. Lastly, the minimum of the candidate times indicates the event time of our hypothetical individual. Note that this flexible approach to simulation from a joint model is completely general and allows for *any* model specification, including higher order polynomials or splines.

Different event rates were chosen by changing the baseline hazard parameters,

θ_0 and θ_1, in the simulations. Table 2.2 and Table 2.3 summarize the simulation results of each of the models with joint analysis and separate analysis of each component. To achieve a 70% event rate with Model II, $\theta_0 = -3$ and $\theta_1 = 0.5$ were used.

	Model I:U_0			Model II:$U_0 + U_1 t$			Model III:$U_0 + U_1 t + U_2 t^2$		
Variable	True Value	Joint(SE)	Sep (SE)	True Value	Joint(SE)	Sep (SE)	True Value	Joint(SE)	Sep (SE)
Longitudinal									
Constant	7	7.000(0.010)	7.000(0.010)	7	7.000(0.031)	7.000(0.031)	7	6.963(0.041)	7.016(0.041)
Continuous	-1	-0.999(0.010)	-1.000(0.010)	-1	-1.001 (0.030)	-1.001(0.030)	-1	-0.991(0.041)	-0.995(0.041)
Binary	0.5	0.499(0.010)	0.499(0.010)	0.5	0.501(0.031)	0.501(0.031)	0.5	0.492(0.041)	0.496(0.041)
Time	0.1	0.100(0.001)	0.100(0.001)	0.1	0.094(0.013)	0.093(0.013)	0.1	-0.249(0.066)	-0.283(0.066)
Survival									
Continuous	0.6	0.601(0.037)	0.595(0.037)	0.6	0.600(0.043)	**0.410(0.042)**	0.6	0.584(0.050)	**0.308(0.042)**
Binary	-0.6	-0.600(0.040)	-0.594(0.040)	-0.6	-0.602(0.042)	**-0.409(0.040)**	-0.6	-0.579(0.049)	**-0.299(0.040)**
Association									
γ_1	0.4	0.404(0.116)		0.4	0.398(0.044)		0.4	0.414(0.047)	
γ_2				0.7	0.701(0.033)		0.7	0.634(0.052)	
γ_3							0.2	0.199(0.013)	
Random Effects									
σ^2_{u0}	0.1	0.100(0.005)	0.100(0.005)	0.9	0.895(0.039)	0.895(0.039)	1.5	1.497(0.083)	1.482(0.083)
σ_{u0u1}				-0.04	-0.040(0.014)	-0.042(0.014)	0.5	0.374(0.085)	0.349(0.083)
σ_{u0u2}							0.5	0.539(0.086)	0.476(0.083)
σ^2_{u1}				0.16	0.159(0.008)	0.159(0.008)	1.5	1.238(0.122)	1.234(0.121)
σ_{u1u2}							0.5	0.433(0.076)	0.392(0.071)
σ^2_{u2}							1.5	1.629(0.119)	1.515(0.113)
Noise									
σ^2_ε	0.01	0.010(0.0003)	0.010(0.0003)	0.01	0.010(0.0003)	0.010(0.0003)	0.4	0.400(0.017)	0.404(0.017)
Log likelihoods									
Combined		-4406(125.700)	-4413(125.700)		-5920(114.100)	-6174(112.000)		-9931.388(127.041)	-11082.600(124.663)
Longitudinal		1048(62.240)	1052(63.280)		-1150(50.930)	-1154(51.730)		-5963.460(162.705)	-5961.603(162.721)
Survival		-5458(90.940)	-5465(90.910)		-4767(93.710)	-5020(91.680)		-3967.928(76.612)	-5121.001(94.567)

Table 2.2: Simulation results based on 1000 sample size and 500 simulations, with a 70% event rate, Scenario I.

	Model I:U_0			Model II:$U_0 + U_1 t$			Model III:$U_0 + U_1 t + U_2 t^2$		
Variable	True Value	Joint(SE)	Sep (SE)	True Value	Joint(SE)	Sep (SE)	True Value	Joint(SE)	Sep (SE)
Longitudinal									
Constant	7	6.999(0.011)	6.999(0.011)	7	7.001(0.032)	7.001(0.032)	7	6.999(0.040)	7.005(0.040)
Continuous	-1	-1.000(0.010)	-1.000(0.010)	-1	-1.000(0.031)	-1.000(0.031)	-1	-1.001(0.041)	-1.000(0.041)
Binary	0.5	0.501(0.010)	0.501(0.010)	0.5	0.501(0.030)	0.501(0.030)	0.5	0.498(0.040)	0.498(0.040)
Time	0.1	0.100(0.001)	0.100(0.001)	0.1	0.098(0.013)	0.098(0.013)	0.1	0.085(0.042)	0.072(0.041)
Survival									
Continuous	0.6	0.604(0.065)	0.601(0.065)	0.6	0.593(0.065)	**0.386(0.063)**	0.6	0.465(0.086)	**0.072(0.067)**
Binary	-0.6	-0.600(0.070)	-0.598(0.069)	-0.6	-0.600(0.071)	**-0.395(0.066)**	-0.6	-0.463(0.087)	**-0.074(0.071)**
Association									
γ_1	0.4	0.388(0.207)		0.4	0.395(0.067)		0.4	0.390(0.073)	
γ_2				0.7	0.697(0.037)		0.7	0.533(0.031)	
γ_3							0.2	0.159(0.007)	
Random Effects									
σ^2_{u0}	0.1	0.100(0.005)	0.100(0.005)	0.9	0.898(0.038)	0.898(0.038)	1.5	1.498(0.077)	1.501(0.077)
σ_{u0u1}				-0.04	-0.039(0.012)	-0.040(0.012)	0.5	0.496(0.062)	0.490(0.062)
σ_{u0u2}							0.5	0.501(0.057)	0.516(0.057)
σ^2_{u1}				0.16	0.159(0.007)	0.159(0.007)	1.5	1.503(0.084)	1.448(0.082)
σ_{u1u2}							0.5	0.502(0.056)	0.481(0.056)
σ^2_{u2}							1.5	1.501(0.070)	1.524(0.071)
Noise									
σ^2_ε	0.01	0.010(0.0002)	0.010(0.0002)	0.01	0.010(0.0002)	0.010(0.0002)	0.4	0.398(0.010)	0.400(0.011)
Log likelihoods									
Combined		1326(149.500)	1320(149.700)		-1893(114.400)	-2100(119.700)		-13091.160(149.516)	-13738.060(165.146)
Longitudinal		3186(79.620)	3187(79.620)		-238.7(60.760)	-238.7(60.760)		-12042.280(140.472)	-12041.900(140.473)
Survival		-1861(104.300)	-1866(104.400)		-1654(95.380)	-1861(100.900)		-1048.879(67.382)	-1696.154(94.251)

Table 2.3: Simulation results based on 1000 sample size and 500 simulations, with a 25% event rate, Scenario II.

The results can be summarized as follows:

- Joint estimates and separate estimates of the longitudinal submodel are all good for Models I and II in both scenarios, since random effects estimates do not change much. Separate estimates are slightly better when the event rate is 70%. Survival estimates of joint model are almost unbiased under Scenario I for Model I and II. On the other hand, separate analysis estimates of Model III under Scenario I are 50% biased; however, they are a long way off. Furthermore, joint model estimates are approximately 25% biased under Scenario II. The reason for this may be the lack of events. One can suggest that quadratic models need more events to be able to capture unbiased parameter estimates. Severe bias occurs for survival parameters, particularly in terms of the underestimation of the survival submodel parameters, when latent association is ignored.

- In terms of association parameters, joint model estimates are slightly biased. However, when the event rate is low, the bias is greater than when the event rate is relatively high. One can therefore suggest that effects such as the number of events and sample size are important to estimate the association parameter.

- All estimates of the variance covariance parameter and noise parameter are satisfactory under the two analyses.

- Log likelihoods dictate that for each of the Models I, II and III, under both scenarios, there is a strong association between repeated measurements and survival times, with large increases in log likelihoods in comparison with the values obtained under separate analyses.

- Overall, results under the joint modelling approach are good, although one can note small bias in survival submodel parameters for smaller samples and more complex association structures. The more complicated the model is, the more underestimation of the results of the parameters ensues under separate analysis.

In summary, if there is no association, there is no benefit of fitting joint model (Henderson et al., 2000) the simulation studies demonstrate that the joint model is better choice over separate analyses of each outcome for the data when the association between longitudinal measurements and survival times is 0.4. The principal advantage of this approach over separate analyses is the correct treatment of noisy and incompletely observed time-varying covariate information, which enables unbiased estimation of the relationship between the two, and accurate estimation of any baseline covariates in the survival submodel. This is particularly important, as binary covariates here is akin to 'treatment' in a real dataset. When there is an association between longitudinal and survival outcomes, joint modelling is more accurate than separate analysis of each component.

2.7.2 Simulation study II

This section investigates the consequences of fitting a misspecified joint model to data that is generated under the joint model with a different latent association structure, and the impact on estimates of parameters. The aim is to evaluate the bias, mean square errors (MSE), and the coverage probability that may occur in the coefficients. The data generation method is the same as explained in the previous section, with a 50% event rate. Misspecification is categorized into two types to evaluate the misspecified models when data is more complex than the model fitted, and when the data is simpler than the model fitted, as shown in Table 2.4 and Table 2.5, respectively.

Results based on 100 simulations, each with a sample size of $n = 250$, are presented in Table 2.4. The reason of performing less simulation than the previous simulation study is the fact that the variance estimator of the each parameter had to be calculated to be able to calculate the coverage probability. The bootstrapping is one way of achieving this. Due to computational burden of the bootstrap method, a smaller sample size of 100 simulations were performed. Data are generated under a random quadratic model, and fitted with both a random intercept model and an intercept-and-slope model, referred to as Sce-

narios A and B, respectively. In Scenario C, data are generated under a random intercept and slope model, and fitted with a random intercept model. These scenarios are presented in Table 2.5. Separate association for each random effect component was also chosen, so γ is not scalar. Overall, when the true model is far from the fitted model, for instance in Scenario A, one can notice that while the bias in the association parameter is small, the bias of the random effects variance and error are considerably high, so both MSEs and the coverage probability (CP) are noticeably small, except the CP for the association parameter, which is in fact mostly zero. However, when the true model is only one step away from the fitted model, for instance in Scenarios B and C, bias in the random effects and error terms are not so high. However, the bias of the associations is high compared to Scenario A. Overall, the coverage probabilities are low, except in associations for Scenarios A and B.

Results based on 200 simulations, each with sample size $n = 250$, are given in Table 2.5. The reason why these simulations are performed more than the simulations in Table 2.4 is that around 30% of bootstrap samples are "not converged". Therefore, more simulations were performed here to be able to collect more reliable results. The scenarios fitted in this table are data generated with the random intercept model and the random intercept and slope model fitting, and data generated with the random intercept and slope model and the random quadratic model fitting, referred to Scenarios D and E, respectively. In addition, there were 15 singularity errors on the variance-covariance matrix in Scenario D. As such, those simulations were discarded, while the rest were taken into account. It can clearly be seen from this table that the bias in the association parameter is extremely high, especially for Scenario D. This may be a result of forcing the model to fit it with a parameter that is redundant in the correctly specified model. Here, the coverage probabilities are still low, except for the association parameters, which are truly zero. An attempt was made to fit the random quadratic model, with the true model being the random intercept model, to investigate the effect of forcing a model very far from the true model. In addition to having many unconverged bootstrap samples, singular-

ity errors were also experienced many times in spite of the fact that a wide range of parameters were attempted. As such, the empirical results show that if the true model is simpler than the fitted model, it is inevitable to observe non-convergent bootstrap samples. Furthermore, if the fitted model is much more complicated than the true model, experiencing singularity errors in the random effects variance covariance matrix becomes possible. Small sample size may have contributed here too, based on the arguments discussed in the previous section.

Simulation study I and II indicate that there is no difference between using MC and bootstrap methods. They are both two powerful techniques to reinforce the data. However, the MC simulation need a hypothesis on the distribution form of the data. The distribution forms of the data here is quite evident.

	Scenario A				Scenario B				Scenario C			
Variable	True Values	Bias	MSE	CP (95%)	True Values	Bias	MSE	CP (95%)	True Values	Bias	MSE	CP (95%)
Longitudinal												
Constant	7	0.658	0.451	0	7	0.113	0.025	0.210	7	0.131	0.040	0.200
Continuous	-1	-0.101	0.038	0.220	-1	-0.006	0.005	0.230	-1	-0.072	0.019	0.220
Binary	0.5	0.109	0.035	0.220	0.5	0.018	0.020	0.280	0.5	0.114	0.074	0.290
Time	0.1	-0.563	0.329	0	0.1	-0.113	0.017	0.170	0.1	-0.189	0.037	0
Survival												
Continuous	0.6	-0.193	0.051	0.250	0.6	-0.132	0.034	0.390	0.6	-0.092	0.023	0.440
Binary	-0.6	0.207	0.054	0.300	-0.6	0.130	0.055	0.450	-0.6	0.121	0.066	0.490
Association												
γ_1	0.4	-0.002	0.004	0.930	0.4	0.201	0.062	0.810	0.4	0.291	0.090	0
γ_2	0.7				0.7	-0.400	0.161	0.110	0.7			
γ_3	0.2				0.2							
Random Effects												
σ^2_{u0}	1.01	4.067	16.959	0	1.010	0.270	0.101	0.300	1	1.913	3.745	0
σ^2_{u1}	0.41				0.410	0.880	0.790	0.210	0.2			
σ^2_{u2}	0.03				0.03							
Noise												
σ^2_ε	0.4	6.214	39.292	0	0.4	0.545	0.304	0	0.4	0.666	0.450	0

Table 2.4: Joint model parameter estimation results under misspecified random effects models. Complicated data with simple model fit. MSE: mean square error, CP: coverage probability.

	Scenario D				Scenario E			
Variable	True Values	Bias	MSE	CP (95%)	True Values	Bias	MSE	CP (95%)
Longitudinal								
Constant	7	0.004	0.094	0.232	7	-0.066	0.173	0.245
Continuous	-1	-0.007	0.040	0.238	-1	0.029	0.102	0.225
Binary	0.500	-0.054	0.168	0.449	0.500	-0.028	0.332	0.525
Time	0.100	-0.002	0.0004	0.238	0.100	-0.008	0.001	0.240
Survival								
Continuous	0.600	0.018	0.017	0.481	0.600	0.024	0.031	0.460
Binary	-0.600	-0.033	0.067	0.492	-0.600	-0.018	0.107	0.490
Association								
γ_1	0.400	0.004	0.002	0.492	0.400	0.003	0.002	0.310
γ_2	0	-107.863	3599506.000	0.989	0.700	0.013	0.026	0.545
γ_3					0	0.682	8.427	0.980
Random Effects								
σ^2_{u0}	9	-0.047	0.782	0.476	20	-0.228	2.998	0.305
σ^2_{u1}	0	0.002	0.00001	0.470	0.100	0.010	0.001	0.320
σ^2_{u2}					0	0.0003	0.00000	0.325
Noise								
σ^2_ε	1	-0.020	0.003	0.892	0.500	-0.005	0.001	0.945

Table 2.5: Joint model parameter estimation results under misspecified random effects models. Simple data with complicated model fit. MSE: mean square error, CP: coverage probability.

2.8 Analysis of liver cirrhosis data

The application of the joint model to the dataset described in Section 1.5.2 is presented in this section. The primary interest of this study is to investigate the effect of treatment on survival times after adjusting for the repeated measurements of prothrombin index on the time to all-cause mortality. Data is available for 488 patients, randomly allocated at diagnosis to prednisone (251) or placebo (237). The measurements were taken at baseline (at entry) and subsequently scheduled to be taken at 3,6 and 12 months, and annually thereafter. The time-to-event outcome was death. More detailed information can be found in Andersen et al. (1993).

Figure 2.2 illustrates the relationship between the trajectories of the repeated measurements (prothrombin index) and the time (years) relative to the event (death) by plotting the observed longitudinal profile of 488 patient against time with a lowess smoother overlaid to give indication of the mean profile, showing that patients who encountered the event had generally lower values of prothrombin index. The timescale is adjusted by taking away the observed event or censoring time. This figure shows a trend in the prothrombin index in the first a few years before the event, possibly indicating that there could be an association between lower values of prothrombin index and an increased risk of encountering the event. The joint model was fitted with three different latent association structures. Table 2.7 presents the parameter estimates, standard errors and the maximized log likelihoods under joint analysis. For an explanation of the choice of covariates, see Henderson et al. (2000). The standard errors are obtained via bootstrap technique and 200 bootstrap samples are used to compute standard errors. For each of the models, there is a joint model of this dataset applied in Henderson et al. (2002) with slightly different model fit. The authors used proportional association between longitudinal and survival data, with an inclusion of interaction term between treatment and baseline time, as well as a stationary Gaussian process latent term. Here, a slightly simpler model was fitted with separate association between each component,

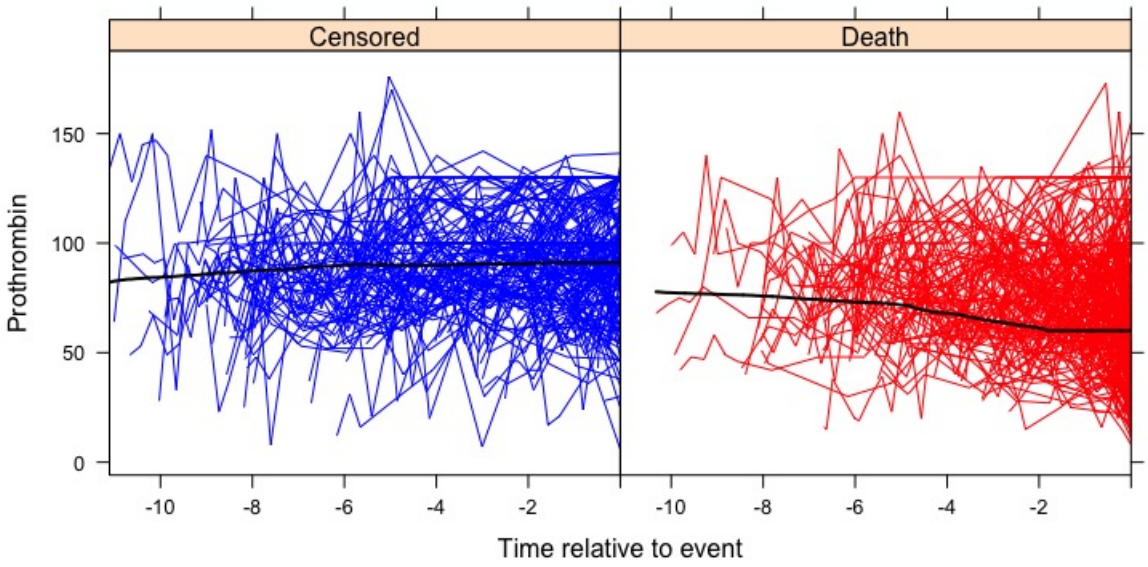

Figure 2.2: Longitudinal trajectories of prothrombin index of each individual censored (left panel) or died (right panel).

without interaction between treatment and baseline time, resulting in obtaining the standard errors through bootstrap. While here the random quadratic model was also fitted, rather than having a stationary Gaussian process (SGP) in the latent association, as our most complex model does, the authors stated that "the widely used random intercept and slope model for longitudinal data introduced by Laird and Ware (1982) is particularly fragile when applied to data consisting of relatively long sequences but with a relatively high dropout rate".

For each of the models, there is strong association between prothrombin index and time to death, with large increases in log likelihood in comparison with the log likelihoods obtained under separate analysis, as indicated in Table 2.6. Negative estimates of association indicate that low prothrombin levels are associated with an increased risk of having the event, as would be clinically expected.

Although it seems from Table 2.7 that the best model is Model III, improved the log likelihoods, one can say that it may not be an optimal model for this

	Longitudinal	Survival	Combined
Model I	-13331.26	-1880.589	-15211.84
Model II	-13273.33	-1880.589	-15153.92
Model III	-13267.62	-1880.589	-15148.21

Table 2.6: Log likelihoods for liver cirrhosis trial based on separate analysis

dataset based on our experience in Section 2.7.2. In that section, the effects of fitting a misspecified joint model were investigated, revealing that when the true model was simpler than the fitted model, there were at least 30% of bootstrap samples which failed to converge. Based on the simulations in Scenarios D and E we can state that number of converged bootstrap samples is another criteria that needs to be considered in terms of fitting best model. During the fitting of Model III, non-convergence was experienced in each bootstrap sample. Therefore, one can say with confidence that Model III is not an optimum model choice for the liver cirrhosis dataset. One can say that Model II is the most convenient model for this dataset comparing the log likelihoods as well as taking into account the convergent number of bootstrap samples. Thus, a recommendation to researchers would be to revise the choice of model if there are a substantial number of bootstrap samples which fail to converge while fitting a joint model.

2.9 Discussion

The joint models of repeated measurements and time-to-event data has been a popular and highly-active research field over the past two decades (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000). A number of excellent reviews have been carried out such as (Sousa, 2011; Tsiatis and Davidian, 2004; Yu et al., 2004). As such, joint modelling has become a well-established methodological framework in terms of association between longitudinal and survival outcomes.

In this chapter, the methodological foundations of the joint modelling approach were presented along with the two requisite building blocks: longitudinal data

	Model I		Model II		Model III	
Variable	Est	SE	Est	SE	Est	SE
Longitudinal						
Constant	76.179	1.284	76.916	1.357	76.849	1.461
Time	1.246	0.329	0.149	0.481	-0.021	0.434
Treatment	-7.203	1.870	-7.030	1.937	-6.952	2.065
Treatment:Time	0.430	0.476	0.264	0.515	0.348	0.535
Survival						
Treatment	0.060	0.139	0.088	0.169	0.098	0.158
Association						
γ_1	-0.039	0.004	-0.037	0.004	-0.038	0.005
γ_2			-0.054	0.012	-0.071	0.014
γ_3					-0.106	0.040
Random Effects						
σ^2_{u0}	379.595	24.440	358.583	27.234	366.471	30.561
σ^2_{u1}			15.795	4.215	43.914	12.752
σ^2_{u2}					0.291	0.166
Noise						
σ^2_{ε}	342.642	14.939	296.725	13.898	289.769	13.984
Log likelihoods						
Combined	-15118.47		-15041.73		-15029.57	
Longitudinal	-13333.22		-13278.4		-13273.17	
Survival	-1785.25		-1763.331		-1756.398	

Table 2.7: Parameter estimates (Est) and standard errors (SE) of liver cirrhosis trial, with three different latent association structure: random intercept model (Model I), random intercept and slope model (Model II), and random quadratic model (Model III).

analysis and time-to-event analysis. The general modelling strategy was based on a specification of two linked Gaussian random processes, $W_1(t)$ and $W_2(t)$, and the association between them. Different kinds of latent association structure were also presented.

Two simulation studies were conducted. The first simulation study investigated the performance of joint modelling approach, and compared it with that of the separate analysis in the presence of separate latent association. Overall, it was concluded that the survival submodel estimates were substantially underestimated in the presence of latent association when it was ignored. The second simulation study investigated the effects of misspecifying a latent association structure on parameter estimates under the shared parameter models framework. The results demonstrated that the coverage probabilities are close to zero under misspecification. Although bias in the parameter estimates seemed low, bias in the association parameter, especially when the model is fitted with an redundant association parameter, was considerably high.

The chapter also presented an analysis of the liver cirrhosis dataset, where the repeated measurements were prothrombin index, and the survival outcome was time-to-death. The results demonstrated the usefulness of prothrombin index as a predictor for the hazard of death. Despite having the highest log-likelihoods in Model III, there was only around 10% converged bootstrap samples, while the rest did not converged. As such, based on the experience gathered from simulation study II, one can conclude that this model is not suitable for this dataset, which may be due to estimating an extra parameter when the true value of the association is zero, as in such a case, the model may be forced to be fitted to an extra unavailable random effect.

The joint modelling considered in this chapter is called the shared random effects model. Another option would be the latent class models (Proust-Lima et al., 2014). While the results obtained here were gathered by using ML estimation, the Bayesian inference represents an alternative way to obtain joint

modelling results (Guo and Carlin, 2004).

Chapter 3

Development of the Methodology for Joint Modelling of Time-To-Event and Multilevel Longitudinal Data with An Application to the SLS Data

3.1 Introduction

In Chapter 2, the main concepts of joint modelling of longitudinal and survival data were introduced, simulation studies were conducted, and the motivation of a joint model analysis was described. In this chapter, a methodological extension to the joint modelling framework is proposed to address the issue of, specifically, a centre-level random effect, although the general idea extends to any hierarchical model.

So far, there has been little discussion about the centre-level effect owing to the lack of motivating applications and available software within joint modelling framework. The statistical challenge is to develop a methodology and software to incorporate multi-level models in the joint modelling framework. In particular, it is of interest to allow the centre level effect in joint model applied for

the SLS data described in Section 1.5.1. Current software allows for individual patient heterogeneity, but precludes the modelling of higher-level variations. In a wider context, multilevel structure could involve geographic areas, types of hospital, or, more acutely, family membership, e.g. twins. Extensions to incorporate multilevel aspect are compromised by increasing dimensionality, and sophisticated solutions will need to be sought in order to keep computational times at a reasonable level.

Many existing approaches to joint modelling link the two responses via subject-level random effects, as mentioned earlier. These models include only subject-level random effects as surrogates for the unobserved process. In contrast, in this chapter, the joint modelling methodology with two levels of nested random effects (subject-level and subjects nested within clusters) is extended through the use of a mixed effects model for the repeated measurements that incorporates both subject- and cluster-level (also called centre-level) random effects.

Ratcliffe et al. (2004) developed a joint model for multi-level repeated measurements and survival outcome. The application of their model was hemoglobin levels (repeated measurements) and patients' survival, clustered within 19 centres. The authors linked the two responses via the common cluster-level random effects, or frailties. They assumed that patients' survival depends only on random effects at the centre-level. Furthermore, they used the EM algorithm to estimate the parameters and concluded that the performance of the cluster-level linkage was better than a subject-level link in the presence of clustered data.

Liu et al. (2008) proposed a comprehensive joint model which assumes that survival of patients depends on random effects at both levels. The authors performed an analysis with a random sample of the U.S. renal data system dataset, including 126 dialysis centres. In addition, they adopted a piecewise log-linear baseline hazard for death, and used the Gaussian quadrature technique implemented in aML (Multiprocess Multilevel Modeling software (Lillard and Panis,

2003)).

The proposed model introduced here is more comprehensive, and the developed software is user-friendly. The modelling concept of Wulfsohn and Tsiatis (1997) is followed, and a joint model for survival, and, now hierarchical (multilevel), longitudinal data is formulated. The proposed model enables making inferences about association between outcomes through the shared random effects at both patient/subject and centre/hospital/cluster levels. A variety of random effect structures are adopted for both subject and centre levels. A noticeable feature of the proposed modelling plan, as in Henderson et al. (2000), is to postulate a latent bivariate Gaussian process $W(t) = \{W_1(t), W_2(t)\}$ for the longitudinal and survival components, respectively. In addition, the longitudinal measurements and survival processes are assumed to be conditionally independent, given the random effects and covariates. Therefore, association is set out via the cross-correlation between $W_1(t)$ and $W_2(t)$. The simplest case is started with, in which the random intercept model is allowed at both the subject and centre-levels.

The joint modelling concept proposed by Wulfsohn and Tsiatis (1997) is further considered in this chapter. The random effect model considered as the most complex structure is random intercept and slope model at subject and centre levels (specified as Model D subsequently) to indicate how the model looks like when there are four components in the random effects model along with separate association structure. However, the estimation method is calculated for the random intercept model at both levels (specified as Model C subsequently sections) along with the proportional association structure for likelihood estimation method, simulations and an application.

The rest of the chapter is structured as follows: Section 3.2 describes the model and notation, while Section 3.3 sets out the assumption that $W_1(t)$ and $W_2(t)$ can be specified through a linear random effects model, and provides the observed data likelihood. The EM algorithm proposed by Wulfsohn and Tsiatis

(1997) is extended by adding a centre level of random effects in the subsequent section. In Section 3.5, simulation studies are performed and the results are discussed. The final section provides a brief summary and discusses possible extensions to the work.

3.2 Model and notation

Let T^*_{hi} denote the *true* event time for the i^{th} individual at centre h, where $h = 1, 2, \cdots, H$, and T_{hi} denote the observed event time. $T_{hi} = \min(T^*_{hi}, C_{hi})$, where C_{hi} is the censoring time. In this study, the event times are assumed to be subject to right censoring, which is a common assumption in most clinical trials and longitudinal studies. Furthermore, $\delta_{hi} = I(T^*_{hi} \leq C_{hi})$ is defined as a failure indicator, which is equal to 1 if the failure time is observed, and 0 otherwise.

In this section, a single time dependent covariate is examined or biomarker (i.e. univariate longitudinal data). Let $Y_{hij}(t)$ denote the observed outcome for patient i ($i = 1, 2, \ldots, n_h$, where n_h is the number of individuals in centre h) at visit j ($j = 1, 2, \ldots, J_i$, where $j = 1$ is baseline, and J_i is the number of measurements for patient i) at centre h ($h = 1, 2, \ldots, H$, where H is the number of centres). The total sample size is given by $n = \sum_{h=1}^{H} n_h$. Furthermore, let $\boldsymbol{Y_{hi}} = (Y_{hi1}, \ldots, Y_{hiJi})'$, and $\boldsymbol{Y_h} = (\boldsymbol{Y'_{h1}}, \ldots, \boldsymbol{Y'_{hn_h}})'$ be the collection of longitudinal measurements for individual i at centre h and the complete set of observations at centre h, respectively.

Each individual may have a different number of measurements (e.g. the i^{th} subject has J_i observations, for $i = 1, 2, \ldots, n_h$). We define $t_{hi} = (t_{hij} : t_{hij} \leq T_{hi})$, where t_{hi} is the measurement time for the corresponding covariate. All individuals start at $t = 0$. The longitudinal measurements at the measured times are $\boldsymbol{Y_{hi}} = (Y_{hij} : t_{hij} \leq T_{hi})$. Along with the longitudinal measurements, the time-independent fixed covariates such as gender, baseline age and treatment group, are considered in this study.

The model is set and the likelihood for the combined data is derived through the common random effects by considering different types of random effects structures. Generally, we define $\boldsymbol{W_k}(t) = \boldsymbol{D_k}(t)'\boldsymbol{U} : k = 1, 2,$, where $\boldsymbol{U}$ is the multivariate Gaussian random effects matrix, and $\boldsymbol{D_k}(t)$ are vectors of potentially time-varying explanatory variables. This approach allows for the situation where the association between the two submodels is set out in terms of a variety of components, such as the random intercept and/or slope for subject-specific time and centre-specific time trend, rather than the value of $\boldsymbol{W_1}(t)$ itself. The sequence of longitudinal observations $(Y_{hi1}, Y_{hi2}, \ldots, Y_{hiJi})$ at times $\boldsymbol{t_{hi}} = (t_{hi1}, t_{hi2}, \ldots, t_{hiJi})$ is measured as follows:

$$\begin{aligned} Y_{hij} &= \boldsymbol{X_{1i}}\boldsymbol{\beta_1} + \boldsymbol{W_{1hi}}(t_{hij}) + \varepsilon_{hij}, & i &= 1, \ldots, n_h \\ \varepsilon_{hij} &\sim N(0, \sigma_\varepsilon^2) & j &= 1, \ldots, J_i \\ \boldsymbol{W_{1hi}}(\boldsymbol{t_{hi}}) &= \boldsymbol{D_{1hi}}(\boldsymbol{t_{hi}})\boldsymbol{U_{hi}} & h &= 1, \ldots, H \end{aligned} \tag{3.1}$$

where $\boldsymbol{W_{1hi}}(\boldsymbol{t_{hi}})$ is the latent process for the i^{th} individual at centre h, and ε_{hij} is a sequence of mutually independent measurement errors. The error is assumed to be independent of the random effects.

The failure hazard model is set through the Cox proportional hazards model, where the hazard depends on the current value of the longitudinal marker. The proportional hazards model, including latent process and fixed covariates, can be expressed as follows:

$$\begin{aligned} h_i(t|\boldsymbol{U_{hi}}, \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{2i}}) &= h_i(t|\boldsymbol{U_{hi}}, \boldsymbol{x_{2i}}) \\ &= h_0(t)\exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + \boldsymbol{W_{2hi}}(\boldsymbol{t})\} \end{aligned} \tag{3.2}$$

where $h_0(t)$ is the baseline hazard, and $h_i(\cdot)$ is hazard function for patient i. Note that (3.1) and (3.2) are conditionally independent given X and the latent processes. The proportionality assumption is applied between W_{2hi} and W_{1hi} as *latent association*, namely $W_{2hi} = \gamma W_{1hi}$, where Wulfsohn and Tsiatis (1997) used this assumption.

The random effect models considered in this chapter are random intercept

only model and random intercept and slope model at both levels. The simplest model is random intercept model at both levels with proportional association structure, the most complex model is random intercept and slope model with separate association structure for each random effect components. As begin with the former is considered for theoretical derivations and simulation studies; however the latter model is also given in order to enlighten the interested readers. They are explained as follows:

- Model C:

$$W_{1ih}(t) = W_{1ih} = u_{0i} + v_{0h} \tag{3.3}$$

where

$$\begin{pmatrix} u_{0i} \\ v_{0h} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & 0 \\ 0 & \sigma^2_{v0} \end{pmatrix}\right]$$

, $i = 1,...,n$ and $h = 1,...,H$.

- Model D:

$$W_{1ih}(t) = (u_{0i} + v_{0h}) + (u_{1i} + v_{1h}) \times t \tag{3.4}$$

where

$$\begin{pmatrix} u_{0i} \\ u_{1i} \\ v_{0h} \\ v_{1h} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & \sigma_{u01} & 0 & 0 \\ \sigma_{u01} & \sigma^2_{u1} & 0 & 0 \\ 0 & 0 & \sigma^2_{v0} & \sigma_{v01} \\ 0 & 0 & \sigma_{v01} & \sigma^2_{v1} \end{pmatrix}\right]$$

, $i = 1,...,n$ and $h = 1,...,H$. and σ_{u01} is the covariance between u_{0i} and u_{1i} (at subject level), and σ_{v01} is the covariance between v_{0i} and v_{1i} (at centre level).

The individual and centre-level random effects are assumed to independent. However, correlation within these levels is allowed. Hence, the variance covariance matrix of Model C is diagonal, and the variance covariance matrix of Model D is block diagonal. Note that the notations introduced here will also be used in the following chapters.

3.3 The linear random effects model and the likelihood function

Wulfsohn and Tsiatis (1997) assumed a Laird-Ware linear random effects model ($\boldsymbol{W_{1i}}(t) = u_{0i} + u_{1i}t$) in conjunction with a proportionality assumption in the survival submodel. Henderson et al. (2000) extended this model to consider the random intercept and slope to have different effects on the event process. The work developed here allows the model to be extended by adding centre level effects for two different latent association structures: the random intercept model, and the random intercept and slope model, at both the subject and centre-levels. Although not pursued here, a quadratic model would proceed in the same way.

3.3.1 The random intercept and slope model at subject and centre-levels

To help formulate ideas, the random intercept and slope model is considered in the multilevel joint modelling structure (referring to this as the *full model*, *Model D*, since it includes all the random effects in the submodels under this model).

$$\begin{aligned} \boldsymbol{W_{1hi}}(\boldsymbol{t_{hi}}) &= u_{0i} + u_{1i}t + v_{0h} + v_{1h}t \\ &= (u_{0i} + v_{0h}) + (u_{1i} + v_{1h})t \end{aligned} \tag{3.5}$$

where (u_{0i}, u_{1i}) and (v_{0h}, v_{1h}) are the random intercept and slope determining the subject-specific and centre-specific random effects, respectively. Independence across random effects at the individual and centre levels is assumed, however, correlation within individuals and centres is allowed, where they are distributed according to a multivariate normal with block diagonal covariance matrix:

$$\begin{pmatrix} u_{0i} \\ u_{1i} \\ v_{0h} \\ v_{1h} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & \sigma_{u01} & 0 & 0 \\ \sigma_{u01} & \sigma^2_{u1} & 0 & 0 \\ 0 & 0 & \sigma^2_{v0} & \sigma_{v01} \\ 0 & 0 & \sigma_{v01} & \sigma^2_{v1} \end{pmatrix}\right]$$

where the covariance between u_0 and u_1 is σ_{u01} and v_0 and v_1 is σ_{v01}, and u and v are independent. In vector notation, $U_{hi} \sim N(\mathbf{0}, \Sigma)$.

To connect the two submodels, $\boldsymbol{W_{2hi}}(t)$ is proposed:

$$\boldsymbol{W_{2hi}}(t) = \gamma_1 u_{0i} + \gamma_2 u_{1i} * t + \gamma_3 v_{0h} + \gamma_4 v_{1h} * t$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ is the vector of latent association parameters. As mentioned earlier, in the extreme case where $\boldsymbol{\gamma} = \mathbf{0}$, joint modelling offers no advantage to a separate modelling approach.

In this section, the random intercept and slope model for both levels is used as a motivating example. However, it is also possible to have various types of alternative models subject to the combination of random effects. The random effects covariance matrix Σ changes as the random effects change, but the general block-diagonal structure remains unchanged.

3.3.2 Likelihood function

For each individual, we have $(T_{hi}, \delta_{hi}, \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{1i}})$ as observed data. The random effects are not observed. The observed data likelihood is:

$$\prod_{h=1}^{H} \prod_{i=1}^{n_h} \left[\int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{J_i} f(Y_{hij} | U_{hi}, \sigma_\varepsilon^2) \right\} f(\boldsymbol{U_{hi}} | \Sigma) f(T_{hi}, \delta_{hi} | \boldsymbol{U_{hi}}, h_0, \boldsymbol{\gamma}, \boldsymbol{x_{2i}}) d\boldsymbol{U_{hi}} \right] \quad (3.6)$$

where:

$$f(Y_{hij} | U_{hi}, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-1/2} \exp\{-(Y_{hij} - \boldsymbol{x_{1i}}\boldsymbol{\beta_1} - \boldsymbol{D_i}\boldsymbol{U_{hi}})^2 / 2\sigma_\varepsilon^2\}$$
$$f(U_{hi} | \Sigma) = (2\pi|\Sigma|)^{-1/2} \exp\{-(U_{hi})' \Sigma^{-1} (U_{hi})/2\}$$

and,

$$f(T_{hi}, \delta_{hi} | \boldsymbol{U_{hi}}, h_0, \boldsymbol{\gamma}, \boldsymbol{x_{2i}}) \quad (3.7)$$
$$= [h_0(T_{hi}) \exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + \boldsymbol{\gamma}\boldsymbol{U_{hi}}\}]^{\delta_{hi}} \exp\left[- \int_0^{T_{hi}} h_0(s) \exp\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + \boldsymbol{\gamma}\boldsymbol{D_i}\boldsymbol{U_{hi}}\} ds \right]$$

Censoring plays a pivotal role in the parameter estimation process. Ignoring censoring may lead to bias in the estimated parameters Wulfsohn and Tsiatis

(1997). In this model, the censoring does not depend on the random effects, which are very common in clinical trials.

The parameters $\boldsymbol{U}$, Σ, σ_ε^2, γ, $\boldsymbol{\beta_1}$, and $\boldsymbol{\beta_2}$, are estimated through the parametric maximum likelihood, and $h_0(s)$ through the nonparametric maximum likelihood method, where $\boldsymbol{\Omega}$ is set to denote this set of parameters. In each case, the estimate depends on the (unobserved) random effects, and hence, in order to estimate these parameters, the EM algorithm is used.

3.4 Estimation method

The EM algorithm is the most widely used parameter estimation method in the joint modelling literature. It iteratively maximizes the likelihood function of the observed data in two steps, namely the expectation step (E-step) and the maximization step (M-step). E-step computes the expected log-likelihood of the complete data, conditional on the available data, and M-step maximizes the expected log-likelihood. The algorithm iterates between E-step and M-step to estimate the parameters until convergence is achieved.

The observed data for each individual is $(T_{hi}, \delta_{hi}, \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{X_{1i}})$, and the random effects for each individual, U_{hi}, are not actually observed. The complete data likelihood is:

$$\prod_{h=1}^{H}\prod_{i=1}^{n_h}\left[\left\{\prod_{j=1}^{J_i} f(Y_{hij}|U_{hi},\sigma_\varepsilon^2)\right\} f(\boldsymbol{U_{hi}}|\Sigma) f(T_{hi},\delta_{hi}|\boldsymbol{U_{hi}},\boldsymbol{h_0},\gamma,\boldsymbol{x_{2i}})\right] \tag{3.8}$$

where the notations are defined in the previous section.

In the M-step, the score equations are solved, as shown in detail in Appendix A. The survival covariates also have no closed form. However, only the association parameters do not have closed-form estimates. In this case, a one-step Newton-Raphson method is applied. The rest of the parameters have closed-form maximum likelihood estimates. All parameters depend on the conditional expectation of some function of U_{hi}, say $E\{g(U_{hi})|T_{hi},\delta_{hi},\boldsymbol{Y_{hi}},\boldsymbol{t_{hi}},\boldsymbol{X_{2i}},\hat{\boldsymbol{\Omega}}\}$. This

expectation constitutes the E-step of the EM algorithm.

The parameter estimates which have closed-form are:

$$\hat{\boldsymbol{U}} = \frac{\sum_{h=1}^{H}\sum_{i=1}^{n_h} E_i(U_{hi})}{n} \tag{3.9}$$

$$\hat{\Sigma} = \sum_{h=1}^{H}\sum_{i=1}^{n_h} E_i\{(\boldsymbol{U_{hi}} - \hat{\boldsymbol{U}})(\boldsymbol{U_{hi}} - \hat{\boldsymbol{U}})'\}/n \tag{3.10}$$

$$\hat{\sigma_\varepsilon^2} = \sum_{h=1}^{H}\sum_{i=1}^{n_h} \frac{\sum_{j=1}^{J_i} E_i(Y_{hij} - \boldsymbol{x_{1i}}\boldsymbol{\beta_1} - \boldsymbol{D_{hi}}\boldsymbol{U_{hi}})^2}{J_i} \tag{3.11}$$

$$\hat{h_0}(s) = \sum_{h=1}^{H}\sum_{i=1}^{n_h} \frac{\delta_{hi} I(T_{hi} = s)}{E_i\{\exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2} + \gamma \boldsymbol{D_{hi}}\boldsymbol{U_{hi}})\} Z_{hi}(s)} \tag{3.12}$$

where $Z_{hi}(s) = I(T_{hi} \geq s)$, $0 < s \leq 1$ is an at-risk indicator.

$$\hat{\boldsymbol{\beta_1}} = (\boldsymbol{x_1'x_1})^{-1}\boldsymbol{x_1'Y_h^*} \tag{3.13}$$

where $\boldsymbol{Y_h^*}$ is the corrected version of $\boldsymbol{Y_h}$ (i.e. the subtracted random effects from $\boldsymbol{Y_h}$).

The association parameter, γ and $\boldsymbol{\beta_2}$ are updated through the one-step Newton-Raphson algorithm with iterations and let denote them as $\gamma^* = (\boldsymbol{\beta_2}, \gamma)$. For example, at the k^{th} iteration, the parameter estimate for γ^* is

$$\hat{\gamma^*}_{\boldsymbol{k}} = \hat{\gamma^*}_{\boldsymbol{k-1}} + \boldsymbol{I}^{-1}_{\hat{\gamma^*}_{\boldsymbol{k-1}}} \boldsymbol{S}_{\hat{\gamma^*}_{\boldsymbol{k-1}}}$$

where $\boldsymbol{S}_{\hat{\gamma^*}_{\boldsymbol{k-1}}}$ is the score for $\gamma^*_{\boldsymbol{k-1}}$, and $\boldsymbol{I}_{\hat{\gamma^*_{k-1}}}$ is the information for $\gamma^*_{\boldsymbol{k-1}}$.

All of the above expressions require the conditional expectation of the random effects in the maximisation step. It is $E\{g(U_{hi})|T_{hi}, \delta_{hi}, \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{2i}}, \hat{\boldsymbol{\Omega}}\}$, simply denoted as $E\{g(U_{hi})\}$ for convenience, where $\hat{\boldsymbol{\Omega}} = (\hat{\boldsymbol{U}}, \hat{\Sigma}, \hat{\sigma}_\varepsilon^2, \hat{\boldsymbol{h_0}}, \hat{\gamma}, \hat{\boldsymbol{\beta_1}}, \hat{\boldsymbol{\beta_2}})$. The conditional density of random effects given the observed data and the current estimate of the parameters is:

$$f(U_{hi}|T_{hi}, \delta_{hi}, \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{2i}}, \hat{\boldsymbol{\Omega}}) = \frac{f(U_{hi}, T_{hi}, \delta_{hi}|\boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{2i}}, \hat{\boldsymbol{\Omega}})}{f(T_{hi}, \delta_{hi}|\boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{2i}}, \hat{\boldsymbol{\Omega}})}$$

$$= \frac{f(T_{hi}, \delta_{hi}|U_{hi}, \hat{\boldsymbol{h}}_{\boldsymbol{0}}, \hat{\gamma}, \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \boldsymbol{x_{2i}}) f(\boldsymbol{U_{hi}}|\boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{1i}}, \hat{\Sigma}, \hat{\sigma}_\varepsilon^2, \hat{\boldsymbol{\beta}}_{\boldsymbol{1}})}{\int_{-\infty}^{\infty} f(T_{hi}, \delta_{hi}|\boldsymbol{U_{hi}}, \hat{\boldsymbol{h}}_{\boldsymbol{0}}, \hat{\gamma}, \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \boldsymbol{x_{2i}}) f(U_{hi}|\boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{1i}}, \hat{\Sigma}, \hat{\sigma}_\varepsilon^2, \hat{\boldsymbol{\beta}}_{\boldsymbol{1}}) dU_{hi}}$$

$$E\{g(U_{hi})\} = \frac{\int_{-\infty}^{\infty} g(U_{hi}) f(T_{hi}, \delta_{hi} | \boldsymbol{U_{hi}}, \hat{\boldsymbol{h}}_{\boldsymbol{0}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \boldsymbol{x_{2i}}) f(\boldsymbol{U_{hi}} | \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{1i}}, \hat{\Sigma}, \hat{\sigma}_{\varepsilon}^2, \hat{\boldsymbol{\beta}}_{\boldsymbol{1}}) dU_{hi}}{\int_{-\infty}^{\infty} f(T_{hi}, \delta_{hi} | \boldsymbol{U_{hi}}, \hat{\boldsymbol{h}}_{\boldsymbol{0}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \boldsymbol{x_{2i}}) f(\boldsymbol{U_{hi}} | \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{1i}}, \hat{\Sigma}, \hat{\sigma}_{\varepsilon}^2, \hat{\boldsymbol{\beta}}_{\boldsymbol{1}}) dU_{hi}} \tag{3.14}$$

The density $f(T_{hi}, \delta_{hi} | \boldsymbol{U_{hi}}, \hat{\boldsymbol{h}}_{\boldsymbol{0}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \boldsymbol{x_{2i}})$ is given in Equation (3.7). The density $f(\boldsymbol{U_{hi}} | \boldsymbol{Y_{hi}}, \boldsymbol{t_{hi}}, \boldsymbol{x_{1i}}, \hat{\Sigma}, \hat{\sigma}_{\varepsilon}^2, \hat{\boldsymbol{\beta}}_{\boldsymbol{1}})$ can be derived from the joint distribution of U_{hi} and $\boldsymbol{Y_{hi}}$, which is a multivariate normal, i.e.,

$$\begin{pmatrix} \mathbf{Y}_{hi} \\ \mathbf{U}_{hi} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{Y_{hi}}} \\ \boldsymbol{\mu}_{\boldsymbol{U_{hi}}} \end{pmatrix}, \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix} \right)$$

where $\boldsymbol{\mu}_{\boldsymbol{Y_{hi}}} = \boldsymbol{x_{1i}}\boldsymbol{\beta_1} + \boldsymbol{U_{hi}}\mathbf{1}_i$, $\boldsymbol{\mu}_{\boldsymbol{U_{hi}}} = \mathbf{0}$ and $\mathbf{1}_i$ and $\mathbf{0}$ denote the unit and zero vectors, respectively, and have the same dimensions as $\boldsymbol{t_{hi}}$. The covariance matrix elements are:

$$\mathbf{W}_{11} = \begin{Bmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \cdots, & \begin{pmatrix} 1 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \begin{pmatrix} 1 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \cdots, & \begin{pmatrix} 1 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{Bmatrix} + \boldsymbol{I_{m_i}} \sigma_{\varepsilon}^2$$

$$\mathbf{W}_{21} = \begin{pmatrix} \sigma_{00} + \sigma_{01} & \cdots & \sigma_{00} + \sigma_{01} \\ \sigma_{01} + \sigma_{11} & \cdots & \sigma_{01} + \sigma_{11} \end{pmatrix}$$

$$\mathbf{W}_{12} = \mathbf{W}_{21}'$$

$$\mathbf{W}_{22} = \Sigma$$

The conditional density of U_{hi}, given $\boldsymbol{Y_{hi}}$, then follows the standard normal distribution. The density is:

$$\mathbf{U}_{hi} | \mathbf{Y}_{hi} \sim N(\mathbf{W}_{21}\mathbf{W}_{11}^{-1}(\mathbf{Y}_{hi} - \boldsymbol{x_{1i}}\boldsymbol{\beta_1} - \boldsymbol{U_{hi}}\mathbf{1}_i), \mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}) \tag{3.15}$$

Introducing a new notation, $\mathbf{U}_{Y_{hi}} = \mathbf{W}_{21}\mathbf{W}_{11}^{-1}(\mathbf{Y}_{hi} - \boldsymbol{x_{1i}}\boldsymbol{\beta_1} - \boldsymbol{U_{hi}}\mathbf{1}_i)$, and $\mathbf{W}_{Y_{hi}} = \mathbf{W}_{22} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12}$, allows one to simplify the expression. Then, in simple terms, the distribution of $\mathbf{U}_{hi} | \mathbf{Y}_{hi}$ is $N(\mathbf{U}_{Y_{hi}}, \mathbf{W}_{Y_{hi}})$. The evaluation of the expectation of $g(U_{hi})$ is conducted through an m-point Gauss-Hermite quadrature formula (Press et al., 1992). Wulfsohn and Tsiatis (1997) used a two-point Gauss-Hermite quadrature by experimenting with various choices of m, and found that they do not vary much for $m > 2$ for all expectations required in all closed-form maximum likelihood estimates from (3.9) to (3.12). Philipson

et al. (2017) chose $m = 3$ for the implementation of the EM algorithm for the random intercept and slope model (3 grid points for two dimensions). Here, two dimensions of random intercept model at subject and centre levels were chosen, and 3-points quadrature was used for implementation.

3.5 Simulation studies

Simulation studies are carried out not only to evaluate the proposed joint model of multi-level repeated measurements and time-to-event outcome, but to also evaluate the effects of the misspecification of the model with a few scenarios in simulation studies I and II.

3.5.1 Simulation study I

In this section, two simulation studies are carried out to investigate the performance of the proposed model, and estimate the Monte Carlo standard errors and bootstrap standard errors. The data generation method is the same in both simulation studies. Even though the data from the random intercept model at both levels is simulated and analysed in this section, the data simulation method for the random intercept and slope model at both levels is also given, so it can be utilised in the subsequent chapters.

Firstly the simulation studies under two settings of the proposed joint modelling with the multilevel random effect structures are carried out. Data was generated from a model subject to the random intercept in both levels and submodels. In both settings, 1000 datasets were generated, with a sample size of $H = 10$ centres, and $n_h = 100$ individuals at each centre.

In the simulation studies, a group of subjects with discrete measurement times were considered, from 0 to 7. The longitudinal measurements were generated according to (3.1), in which there is one continuous covariate (x_1) and one binary indicator (x_2), along with the indicator of observation times for the fixed covariates. The longitudinal data for all possible follow-up times are first sim-

ulated, using random draws for the multivariate Gaussian random effects, and utilising `mvnorm()` in the `MASS` package (Venables and Ripley, 2002) and residual error terms. Here, the model for latent association is assumed to be either random intercept only, or random intercept and slope at both levels.

The order of the $\boldsymbol{\beta_1}$ vector in each row is intercept, the measurement time, a binary covariate, and a continuous covariate (i.e. $\boldsymbol{\beta_1} = (\mathbf{1}, \boldsymbol{t}, \boldsymbol{x_2}, \boldsymbol{x_1})$). The continuous covariate has standard normal distribution, the binary covariate is chosen from Binomial distribution with 0.5 probability, and the maximum number of measurement times, τ, is set as 8 (i.e. it is truncated at 8).

The `simjoint()` function was modified, which only simulates data from a subject-level random effects to be able to simulate data from subject and centre-levels random effects. The key difference between the original and modified function is that the modified function simulates multi-level longitudinal data for all possible follow-up times using random draws for the multivariate Gaussian random effects and for the centre-level effects. The code to generate this is referred as `simjointml()` and can be found in Appendix A.4.

The survival times were generated according to (3.2). The continuous variable (x_1) and the binary indicator (x_2) were considered once again for each subject with a corresponding regression coefficient vector $\boldsymbol{\beta_2}$. Although the continuous and the binary covariates are the same as in the longitudinal submodel, the corresponding regression coefficients are allowed to be different, as they are typically different in practice. The covariates themselves can, in practice, be different between the longitudinal and survival submodels.

Censoring times are randomly simulated with a scale parameter for an exponential distribution. Censoring was adopted throughout the simulation study, with a specified parameter for the censoring rate of $\exp\{-3\}$, which gives around 70% censored data. There are no tied survival times in the simulated survival times. As such, the assumption concerning the tied events (i.e. having

events with exactly the same event times) would be violated, and the results of Cox model estimation would be directly affected (Borucka, 2014). This can, in practice, be mitigated by adding a small increment to each failure time (i.e. $\varepsilon \sim U(0, 0.01)$) depending on the time scale of the study.

For a random intercept model, regardless of level, the baseline hazard function is specified to be an exponential distribution with

$$\lambda_0(t) = \exp\{\theta_0\}.$$

For a random intercept and slope model, the baseline hazard function is specified to be a Gompertz distribution with

$$\lambda_0(t) = \exp\{\theta_0 + \theta_1 t\}$$

where θ_1 is the shape parameter, and the scale parameter is $\exp\{\theta_0\}$. Here the simulation of the survival times is conditional on a time-dependent process. Two different scenarios are allowed for the simulation of the failure times: a 70% event rate with positive and negative association parameter.

The survival function of the Cox PH model in a multi-level random intercept model is:

$$S(t|U_{hi}, \boldsymbol{x_{2i}}) = \exp\left[-H_0(t)\exp\{\boldsymbol{x_{2i}\beta_2} + \boldsymbol{\gamma D_{hi} U_{hi}}\}\right] \tag{3.16}$$

where $U_{hi} = u_{h0i} + v_{h0}$, and:

$$H_0(t) = \int_0^t h_0(s)ds$$

The cumulative distribution function of the Cox PH model is:

$$\begin{aligned} F(t|U_{hi}, \boldsymbol{x_{2i}}) &= 1 - S(t|U_{hi}, \boldsymbol{x_{2i}}) \\ &= 1 - \exp\left[-H_0(t)\exp\{\boldsymbol{x_{2i}\beta_2} + \boldsymbol{\gamma D_{hi} U_{hi}}\}\right] \end{aligned} \tag{3.17}$$

Let Y' be a random variable having the distribution function of F, and $V = F(Y') \sim U(0,1)$. Furthermore, if a random variable, V, has uniform distribution, then $(1 - V)$ has uniform distribution too (Mood et al., 1974). Let T denote the survival time of the Cox PH model. Then:

$$V = \exp\left[-H_0(t)\exp\{\boldsymbol{x_{2i}\beta_2} + \boldsymbol{\gamma D_{hi} U_{hi}}\}\right] \sim U(\mathbf{0}, \mathbf{1})$$

Subsequently leading to:

$$T = H_0^{-1}\left[\frac{-\log V}{\exp\{\boldsymbol{x_{2i}\beta_2} + \boldsymbol{\gamma D_{hi} U_{hi}}\}}\right] \tag{3.18}$$

where $V \sim U(0,1)$.

	Parameter	True Value	Estimates (SE)	True Value	Estimates (SE)
Longitudinal					
	Intercept	7	7.000(0.033)	7	7.000(0.034)
	Time	-1	-1.000(0.000)	-1	-1.000(0.002)
	Binary	0.5	0.499(0.009)	0.5	0.500(0.009)
	Continuous	0.1	0.100(0.001)	0.1	0.100(0.001)
Survival					
	Continuous	0.4	0.400(0.017)	0.4	0.400(0.018)
	Binary	-5	-5.021(0.175)	-5	-5.014(0.181)
Association					
	γ	0.20	0.207(0.329)	-0.20	-0.186(0.338)
Variance					
	σ^2_{u0}	0.01	0.01(0.001)	0.01	0.01(0.001)
	σ^2_{v0}	0.01	0.009(0.004)	0.01	0.009(0.004)
	σ^2_{ϵ}	0.01	0.01(0.000)	0.01	0.01(0.000)

Table 3.1: Simulation results for the multi-level joint model (based on a sample size of $n = 1000$ and 1000 simulations)

Datasets were generated and analysed with the multilevel joint modelling based on positive and negative association parameters in Table 3.1 with approximately 70% event rate, while datasets were generated and analysed with the multilevel joint modelling based on high and low event rate in Table 3.2. The results can be summarised as follows:

- The results are obtained with the Monte Carlo method in Table 3.1, with 1000 subjects ($H = 10$ and $n_h = 100$) in each simulation, and the simulation is repeated 1000 times. The results are obtained with bootstrap sampling technique in Table 3.2, with a total of 500 subjects ($H = 20$ and $n_h = 25$) and 100 bootstrap samples in each simulation, and the simulation is repeated 100 times.

- In the first table, the two scenarios investigate whether there is any effect on the performance of the model having positive or negative association

		Scenario A		**Scenario B**	
	True Values	Est	SE	Est	SE
Longitudinal					
Constant	7	7.002	0.001	6.998	0.001
Time	-1	-1.000	0.00000	-1.000	0.00000
Binary	0.500	0.496	0.001	0.503	0.001
Continuous	0.100	0.104	0.0003	0.101	0.0003
Survival					
Continuous	0.600	0.612	0.004	0.595	0.009
Binary	-0.600	-0.630	0.012	-0.605	0.038
Association					
γ	0.300	0.318	0.018	0.328	0.052
Random effects					
σ_u	0.400	0.388	0.0002	0.389	0.0002
σ_v	0.120	0.140	0.0004	0.141	0.0004
Noise					
σ_ε^2	0.010	0.010	0.00000	0.010	0.00000

Table 3.2: Simulation results of two scenarios (Scenario A: mean of event rate is 72%, and Scenario B: mean of event rate is 24%). Standard errors are obtained via bootstrap samples.

parameter, while the second table investigates if there is any effect on the performance of the model having high or low event rate (Scenarios A and B, with approximately 72% event rate and 24% event rate, respectively).

- All estimates of the regression coefficients and variance parameters are almost unbiased in the submodels. However, there is negligible bias in association parameters in both scenarios in Table 3.1. The variation was chosen to be small in order to be able to clearly visualise the data and how the association parameter behaves.

- Overall, the results under the joint modelling method with the hierarchical random effects structure show that the methodology and implementation in R have been successful. The performance of the models in terms of association parameters are almost the same, regardless of the repetition of simulations and positive or negative association. The association parameter in each scenario can have very small bias, and the bias can be related to the event rate. Consequently, convincing results were obtained.

3.5.2 Simulation study II

In this section, the effects of misspecification the random effects on the parameter estimates are investigated. The aim here is to explore what happens if the centre-level random effects are ignored when they are needed in the model. The misspecified models are assessed in terms of bias, mean square error (MSE), and coverage probability (CP). Joint data with random intercept and slope model at both levels is first generated, and the random intercept model is fitted at both levels. Hence, the random slopes terms at both levels are ignored. Secondly, the joint data with the random intercept model is generated at both levels, and the random intercept model is fitted at subject-level. Lastly, joint data with random intercept and slope model at both levels is generated, and the random intercept and slope model is fitted at subject-level. These three scenarios are referred to here as Scenarios A, B, and C, respectively. In each setting, 25 subjects ($m = 25$) are generated in each of 20 centres ($H = 20$), resulting in total of $n = 500$ subjects. Each simulation is repeated 100 times, with around 50% event rate.

As it can be seen from Table 3.3, while separate association was applied between longitudinal and survival outcome in Scenarios B and C, proportional association was applied between them in Scenario A, because our code, at the time, only allowed for proportional association. Under Scenario B, 98 simulations out of 100 converged. The other two, which have not been converged, had extremely high standard errors, and as such, were excluded from these simulation results, as they are outliers and would affect the results considerably.

	Scenario A				Scenario B				Scenario C			
Variable	True Values	Bias	MSE	CP	True Values	Bias	MSE	CP	True Values	Bias	MSE	CP
Longitudinal												
Constant	7	-0.196	0.060	0.120	7	0.002	0.001	0.194	7	0.006	0.010	0.190
Time	-1	-0.136	0.024	0.060	-1	0.0003	0.00002	0.235	-1	-0.011	0.005	0.100
Binary	0.500	-0.060	0.030	0.330	0.500	-0.002	0.001	0.194	0.500	-0.011	0.005	0.250
Continuous	0.100	0.032	0.007	0.230	0.100	0.001	0.0001	0.245	0.100	0.001	0.002	0.230
Survival												
Continuous	0.600	0.003	0.007	0.490	0.600	0.009	0.006	0.449	0.600	-0.099	0.015	0.350
Binary	-0.600	-0.016	0.025	0.500	-0.600	-0.024	0.020	0.459	-0.600	0.083	0.031	0.410
Association												
γ_1	0.300	0.022	0.001	0.920	0.400	0.192	0.891	0.949	0.400	0.025	0.026	0.640
γ_2	0.300								0.700	-0.292	0.089	0.390
γ_3	0.300	0.022	0.001	0.920	0.700				0.200			
γ_4	0.300								0.100			
Random Effects												
σ_{u0}	0.548	0.738	0.548	0	0.100	-0.081	0.007	0	0.548	0.199	0.045	0.020
σ_{u1}	0.447								0.447	0.091	0.010	0.330
σ_{v0}	0.447	0.366	0.155	0.070	0.100				0.447			
σ_{v1}	0.316								0.316			
Noise												
σ^2_ε	0.100	1.154	1.355	0	0.400	-0.241	0.058	0	0.400	0.0002	0.0002	0.950

Table 3.3: .

The true model is more complicated than the fitted model. MSE: mean square error, CP: coverage probability. Scenario A: the true model is a random intercept and slope model at both level, and fitted model is a random intercept model at both levels. Scenario B: the true model is a random intercept model at both level, and the fitted model is the same model fitted at individual levels. Scenario C: the true model is a random intercept and slope model at both level, and the fitted model is the same model but fitted at individual levels.

Overall, although bias and MSE seem low in parameter estimates for all scenarios, coverage probabilities are low too. However, the coverage probabilities in survival submodel parameter estimates are higher than those in the longitudinal submodels. In all cases, the coverage probability is low, zero sometimes. The reason may be that the standard errors of the bootstrap samples are small, and confidence intervals are small, and as such, the coverage probabilities are small. It can be said that if the higher level is ignored, and the joint model is fitted only at individual level, one can obtain the association parameter with slightly biased and sufficient coverage probability. However, sufficient coverage probability for the rest of the parameters under the misspecified models cannot be obtained. Hence, centre level effects, when present, can substantially alter the interpretation of parameters.

An important point from Table 3.3 is that ignoring centre level effect at baseline, in Scenario B, causes higher coverage than those simulations ignoring the centre level intercepts and slopes on random effects model. The reason for this is that association parameter in Scenario B is underestimated at 36 times out of 98 simulations and overestimated at the rest 62 simulations with the very high variability. However, in Scenario C, the variability of the association parameter estimates are quite low, that is why the coverage probabilty is low. The magnitude of the chosen parameters for associations (0.7 in Scenario B, 0.2 and 0.1 in Scenario C) can also affect the coverages in these scenairos.

3.6 Application: the scleroderma lung study

We now return to the scleroderma lung study introduced in Section 1.5.1. After the application of the criteria defined in Section 1.5.1, data is available for $n = 136$ patients. The primary outcomes are forced vital capacity (FVC, as % predicted) and treatment failure or death, along with a number of variables recorded at entry. Measurements were obtained at entry, and then scheduled with 3-months intervals during a maximum of two year follow-up time.

Figure 1.1 represents the longitudinal trajectories of the FVC index for surviving patients in each of the treatment groups, truncated at 2 years follow-up. In both groups, data seems noisy and (smoothed) mean profiles have no big changes. Figure 1.2 indicates that more than 70% of patients survive after 2 years from entry, with slightly better prognosis for placebo-treated patients than for those given CYC treatment. Figure 3.1 depicts the joint plot of the SLS data. This plot illustrates the longitudinal profile of each unit, with the last longitudinal measurement prior to event-time taken as the end-point. Elashoff et al. (2016) reported from the preliminary analysis that the change in %FVC over time was highly influenced by its baseline value and maximum lung fibrosis (MAXFIB) as also can be seen from the plot. This model is previously fitted with the same covariates in Elashoff et al. (2016). The authors reported that parameter estimates in longitudinal model for %FVC, and %FVC is associated with its baseline measurement and maximum fibrosis score. Moreover, they stated that the CYC effect appears to be modified by baseline %FVC measurement and maximum fibrosis score. We model the SLS data in this chapter by taking into account for the centre level effect as well as subject level effect in conjunction with a proportionality association.

Parameter estimates, and their standard errors, which are obtained under a joint model with random intercept at both subject and centre levels (Model C), and ignoring the centre level effect are given in Table 3.4. The method used for parameter estimation is explained earlier in this chapter. The standard errors are obtained by the bootstrap sampling method, where number of bootstrap samples is 50, due to the small sample size of this dataset. In this model, there is strong association between %FVC levels and survival times, with the negative association confirming that high %FVC levels are associated with reduced risk, as hypothesised by clinicians in the field. Adding extra variation with the centre level effect improves the model fit as can be seen from the loglikelihoods.

	Model C		**Model A**	
	$W_1(t) = u_0 + v_0$, $W_2(t) = \gamma(u_0 + v_0)$		$W_1(t) = u_0$, $W_2(t) = \gamma(u_0)$	
	Est	SE	Est	SE
Longitudinal				
Constant	-2.128	40.476	-4.855	5.070
MAXFIB	0.029	0.533	0.371	0.743
CYC	10.852	55.126	13.010	7.771
months	0.025	0.004	0.031	0.062
FVC0	1.031	0.005	1.040	0.066
MAXFIB:CYC	-1.416	0.971	-1.722	0.942
CYC:months	-0.132	0.006	-0.139	0.092
CYC:FVC0	-0.108	0.009	-0.131	0.104
Survival				
MAXFIB	0.330	0.041	0.320	0.211
CYC	-0.086	0.223	-0.061	0.370
Association				
γ_1	-0.059	0.002	-0.088	0.083
Random effects				
σ^2_{u1}	4.619	2.632	20.337	3.267
σ^2_{u2}	0.454	0.038		
Noise				
σ^2_ε	34.777	45.683	34.750	5.569
Log likelihoods				
Longitudinal	-3486.376		-3502.201	
Survival	-109.342		-117.527	
Combined	-3595.718		-3619.728	

Table 3.4: The SLS data results for two different $W_1(t)$ models: a random intercept model at both levels with proportional association (Model C), and a random intercept model at subject-level (Model A).

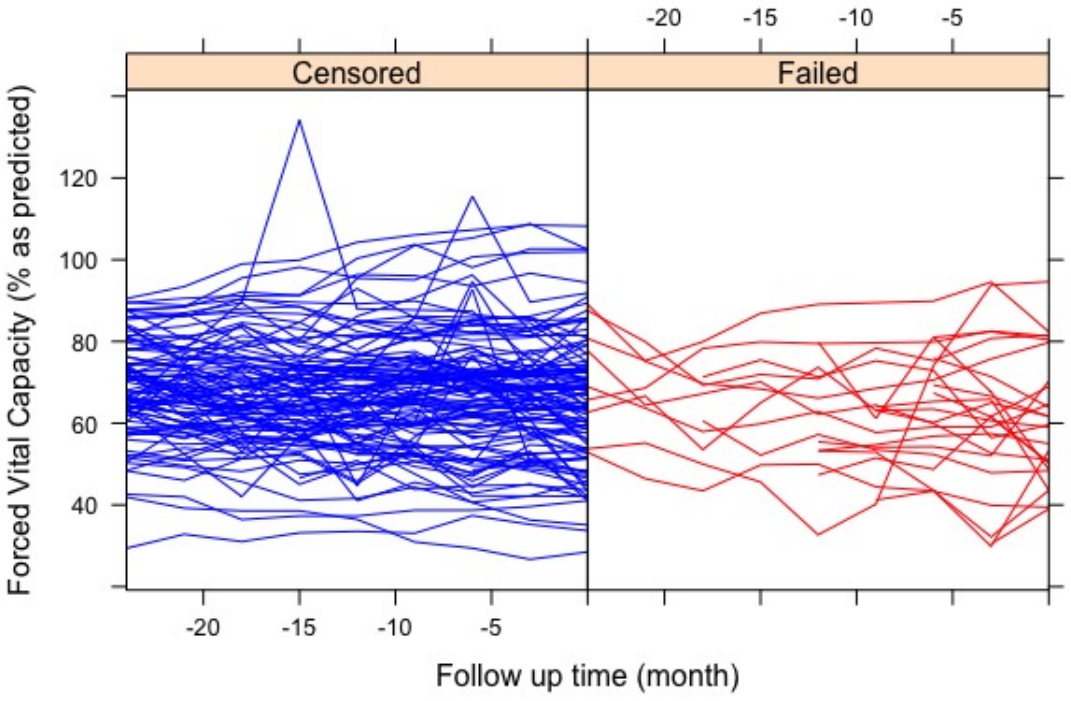

Figure 3.1: Joint plot of the SLS data

3.7 Discussion

As discussed earlier, joint modelling with centre level effect has yet to receive requisite attention. Such situations have data encompassing multiple levels (i.e. patients within a hospital within a region). Ignorance of a centre heterogeneity may lead to significant bias, where the centre level may represent common environmental and socioeconomic status, alongside, or instead of, any formal hierarchy, such as a hospital. This chapter allows one to deal with joint models, with a solitary time-to-event and multi-centre longitudinal measures, where multiple individuals are recruited at each centre and treated by the same clinical site at which they are recruited, thus avoiding to move between centres.

A novel joint model was proposed, consisting of a submodel for multi-level repeated measures, and a Cox submodel for event times. We assumed that the survival time can be correlated with longitudinal measurements at both centre and subject levels, and that variances of clusters are independent. The method

used in Wulfsohn and Tsiatis (1997) was extended to adopt this extended model and use the Gaussian quadrature technique with the code developed in R. To begin with, simulation studies were conducted demonstrating that the estimation method yields satisfactory results. The effects of misspecifying models were then investigated. Furthermore, this method was applied to a real dataset, the SLS data. The application results show that there is strong association between longitudinal measurements and survival outcomes. This method has the capability to accommodate hierarchically structured data to investigate the relationship between longitudinal and time-to-event data in the most appropriate manner.

The simulation studies conducted to assess the performance of the proposed joint model raised an important issue. While this method has negligible bias on estimates of the association parameters, and the rest of the parameters of interest are unbiased according to the first simulation study, Simulation Study II shows that although the bias in the parameters of interest seems quite low (interest may differ for another study), the coverage probabilities are also low. Therefore, it is always important to check bias, MSE and coverage probabilities simultaneously. In this case, the standard errors provided by the bootstrap sampling are quite small, and this, in turn, makes the confidence intervals and coverage probabilities rather narrow.

The proposed method for joint modelling has great flexibility for extensions. For example, it could allow for the separate association effects between longitudinal and survival outcomes, or the model can be more complicated, such as a random intercept and slope model at both levels (as in equation (3.19)), or a random intercept model for the centre level and a random intercept and slope model at the subject level (as in (3.20)).

$$U_{hi} = u_{h0i} + v_{h0} + (u_{h1i} + v_{h1})t \tag{3.19}$$

$$U_{hi} = u_{h0i} + v_{h0} + u_{h1i}t \tag{3.20}$$

where the notations have been specified previously.

One important restriction of the proposed model is that the Gaussian quadrature points must be at least equal to the number of random effects components. Wulfsohn and Tsiatis (1997) claimed that the small number of quadrature points (i.e. $m = 2$) is sufficient. This appealing property of applying the Gaussian quadrature is compromised in case of a rich random effect structure, where one is forced to have at least a minimum number of random effects components (i.e. two in our case, referring to random intercepts at both levels). The model structure then requires evaluation of expressions with a quadruple integral, which are approximated by summations under the quadrature approach. The benefit of quadrature points is diminished in the presence of a lot of random effects. Philipson et al. (2017) implemented a joint model for a random intercept and slope model with $m = 3$ quadrature points. In our work, a 3-point Gauss-Hermite quadrature was adopted in the code used to implement the multilevel joint model for random intercept at both levels. The estimation of parameters in the M-step of the EM algorithm can be evaluated by using a fully adaptive Gauss-Hermite quadrature method, where transformation to independent variables is necessary. Furthermore, evaluation of the expression shown in (A.8) using a 3-point quadrature investigates all possible combinations of 9 (3^2) elements, for two dimensional random components, as in Philipson et al. (2017) and this implementation. All of the combinations of abscissa values across all random effects are indeed desired. For example, the Gaussian quadrature points must be set to at least 4 for the random intercept and slope model at both levels. This means that the length of the transformed independent variables is $4^4 = 256$. This could be a limitation in using Gauss-Hermite quadrature for multilevel models that are subject to demonstrably nonlinear longitudinal profiles at one or more levels. More detailed information regarding the Gauss-Hermite quadrature can be found in Appendix A.2.

Chapter 4

A score test for univariate joint models

4.1 Introduction

In Chapter 3, a methodological extension to the joint modelling framework was proposed to address the issue of centre-level random effects, alongside the common subject-level random effects. The aim was to reduce the unexplained variation and thus have a better fit by considering an extra level, such as families, schools, or hospitals in a district. This could be extended to further levels of structure, although model-fitting using the EM algorithm may become prohibitively costly.

In this chapter, a score test derived for association between longitudinal measurements and survival times by by Henderson et al. (2002) is given and it is implemented with simulation studies for four different random effect structures at both individual and centre-levels. The score test only requires fitting of common models. Through simulations, the power of the score test is investigated under a variety of scenarios, such as different event rates and subject variability. A comparison is also made between martingale and bootstrap variance estimators to provide some indication of the power to detect the association parameter. The former is a theoretical result that holds for large samples, whereas the latter is a computationally intensive empirical approach. More de-

tail on this is given in Section 4.3.

A score test for the association between longitudinal measurements and survival time was developed by Henderson et al. (2002). The authors conducted several simulation studies to check the adequacy of the asymptotic approximations to the null distribution of the proposed test. In this work, the underlying association between longitudinal measurements and survival times was characterised by individual-level deviation of the longitudinal profile from the population mean - through a random intercept and stationary Gaussian process - and the authors derived several pseudo-R^2 statistics to quantify the accuracy of a biomarker for a clinical endpoint; an idea not pursued further here.

In the context of score tests, extensions to the survival sub-model were developed in a series of papers by Ko. Score tests based on joint modelling of repeated measurements and competing risk failure time data (Ko, 2014a) and (Ko, 2014b), survival time allowing individual frailty (Ko, 2010), accelerated failure time (Ko, 2014c), multivariate survival time (Ko, 2017), and the accelerated failure time in multivariate survival data (Ko, 2016) were all derived. In each, simulation studies were carried out in order to examine the empirical type-I error rates of the score test, that is, the power under H_0. Furthermore, Ko used the Martingale variance estimator for calculation of the score statistics.

In the context of latent class joint models, Jacqmin-Gadda et al. (2010) also derived a score test for the null hypothesis of independence between the marker and the outcome given the latent classes, and performed a simulation study to compare the behaviour of the score test to other previously proposed tests including situations where the alternative hypothesis or the baseline risk function are misspecified. It was shown that the score test had theoretical validity and was simple to compute, as computations were performed under the null hypothesis, which typically involves fitting routine models with readily available software. The authors concluded that the score test is a more powerful test than the three proposed tests when the association parameter is close to zero or

a specific value.

A score test is a convenient shortcut to investigate whether one needs a more complicated model by fitting already available routine models. In this chapter, a score test is derived to determine whether a longitudinal marker is associated with the surviving situation under two different random effect structures, at both individual and centre-levels. Several simulation studies are conducted in order to give some indication of the power to detect the association between repeated measurements and time-to-event, and make a comparison between two candidate estimators of the score test statistics: theoretical variance and the bootstrap variance estimators mentioned earlier.

The rest of the chapter is organized as follows: Section 4.2 describes the model and notation, while in Section 4.3, a description of the score test for association is provided. The subsequent section then explains the simulation study in detail and presents the results. The application to the scleroderma lung study is considered in detail in Section 4.5, while Section 4.6 provides some final remarks in close the chapter.

4.2 Model and notation

Following Henderson et al. (2002), the counting process notations in this section are introduced in this section, where these notations are used throughout this thesis for each type of score test considered.

Supposing that we have longitudinal measurements and, possibly censored, event times for n subjects followed over an interval $[0, \tau)$. Let $\boldsymbol{N_i(t)}$ denote the counting process for the i^{th} individual at time t, so that $\{\boldsymbol{N_i(t)} : 0 \leq t \leq \tau\}$, jumping from value 0 to value 1. Furthermore, suppose the i^{th} individual had the event at time t, implying that the counting process function for this subject is equal to zero until time t, and jumps to value 1 at exactly time t. It is then recorded as value 1 until the maximum follow-up time, since only a maximum

of one event is considered here. If the subject is censored, then the counting process for this subject is equal to 0 for all values of t. Let $\boldsymbol{H_i(t)}$ denote the at-risk process. It shows that if the subject is still being followed at time t, that is, it is at-risk for the event. The value of $\boldsymbol{H_i(t)}$ for this subject jumps from 1 to 0 when the follow-up ends, due to experiencing the event or being censored.

Longitudinal measurements, $(y_{i1}, y_{i2}, \cdots,)$ are obtained sequentially at times $(t_{i1}, t_{i2}, \cdots,)$ and we allow the possibility of different number of measurements and dropout any time within the follow-up period for each individual. The longitudinal submodel is then expressed as follows:

$$\begin{aligned} Y_{ij} &= \boldsymbol{x_{1i}}(t_{ij})'\boldsymbol{\beta_1} + W_{1i}(t_{ij}) + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \sigma_\varepsilon^2) \end{aligned} \tag{4.1}$$

where $\boldsymbol{x_{1i}}(t_{ij})'$ is a p_1-vector of baseline covariates for the fixed-effects for individual i, $\boldsymbol{\beta_1}$ is the corresponding vector of regression coefficients, W_{1i} is unobserved latent process, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is a sequence of mutually independent measurement errors. Survival time is associated - typically via proportional association - with the longitudinal measurements through the latent process W_{1i}.

In general, a semiparametric proportional hazards model is assumed, with intensity process at time t, which is given by the semi-parametric multiplicative model:

$$\boldsymbol{\lambda_i}(t) = \boldsymbol{H_i}(t)\alpha_0(t)\exp\{\boldsymbol{x_{2i}}(t)'\boldsymbol{\beta_2} + W_{2i}(t)\} \tag{4.2}$$

where $\boldsymbol{H_i}(t)$ is a predictable 0-1 at-risk process for individual i, $\alpha_0(t)$ is an unspecified baseline hazard, and $\boldsymbol{x_{2i}}(t)$ is the p_2-vector of explanatory variables, $\boldsymbol{\beta_2}$ is the vector corresponding regression coefficients, $W_{2i}(t)$ is the unobserved latent process at time t, and, typically, $W_{2i}(t) = \gamma W_{1i}(t)$. This assumption of proportional association will be relaxed subsequently. Non-informative right censoring of survival times is allowed. Overall, longitudinal responses and survival times are represented using the subscripts Y and T, respectively.

Random intercept and random intercept and slope models are considered at

both subject and centre-levels, assuming the general model (4.1). As such, four different latent processes are specified for the assessment of the power of the score test, of which will be explained more soon.

In this chapter, those four types of latent processes are applied in both simulation studies and a real data application. They are the random intercept model at individual level, the random intercept and slope model at individual level, the random intercept model at individual and centre-levels, and the random intercept and slope model at both levels. They are referred to as Models A, B, C and D, respectively. Models A and C do not include random slope terms, and as such, these models have latent processes independent on time. Other models within this structure are possible, but this is considered to be a reasonable representation of the type of models that are encountered in practice.

The structures of the latent processes in the longitudinal submodels are expressed as follows:

- Model A:

$$W_{1i}(t) = W_{1i} = u_{0i} \tag{4.3}$$

  where $u_0 \sim N(0, \sigma^2_{u0})$, and $i = 1, ..., n$.

- Model B:

$$W_{1i}(t) = u_{0i} + u_{1i} \times t \tag{4.4}$$

  where

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & \sigma_{u01} \\ \sigma_{u01} & \sigma^2_{u1} \end{pmatrix}\right]$$

  and σ_{u01} is the covariance between u_{0i} and u_{1i}, and $i = 1, ..., n$.

Models C and D are the same as given in Section 3.2.

4.3 Score test for association

The score test for association derived by Henderson et al. (2002) is given in this section. The four different random effect model, Models A, B, C and D are considered as the random effect model. This provides the building blocks for us

to extend and develop the methodology to both the multilevel and multivariate models considered later in the thesis.

The score test is based on the separate analysis of the two components (longitudinal measurements and event times) under the null hypothesis ($H_0 : \gamma = 0$). A Gaussian linear model for the longitudinal responses, Y, and a semiparametric proportional hazards model at time t are assumed.

Let τ be the maximum follow-up time. The conditional likelihood of the event history data is:

$$L_\gamma = \left(\prod_t \prod_i (e^{\boldsymbol{x}_{2i}(t)'\boldsymbol{\beta_2} + \gamma W_{1i}(t)} dA_0(t))^{\Delta \boldsymbol{N_i}(t)} \right) \exp\left(- \int_0^\tau \sum_{i=1}^n \boldsymbol{H_i}(t) e^{\boldsymbol{x}_{2i}(t)'\boldsymbol{\beta_2} + \gamma W_{1i}(t)} dA_0(t) \right) \tag{4.5}$$

where $\boldsymbol{N_i}$ is the counting process for the i^{th} individual, and A_0 is the cumulative baseline hazard.

$$A_0(u) = \int_0^u \frac{\boldsymbol{J}(s)}{\sum_{i=1}^n \boldsymbol{H_i}(s) e^{\boldsymbol{x}_{2i}(s)'\boldsymbol{\beta_2}}} d\boldsymbol{N}(s),$$

where $\boldsymbol{N}(s) = \sum \boldsymbol{N_i}(s)$ and $\boldsymbol{J}(s) = \boldsymbol{I}[\sum \boldsymbol{H_i}(s) > 0]$.

Let the unknown parameters be $(\boldsymbol{\theta}, \gamma, \boldsymbol{\beta_2}, A_0)$, where $\boldsymbol{\theta}$ represents all the parameters involved in the linear mixed model. Furthermore, let:

$$U_\gamma(\tau) = \sum_{i=1}^n \left\{ \int_0^\tau W_{1i}(t) d\boldsymbol{N_i}(t) - \int_0^\tau W_{1i}(t) \boldsymbol{H_i}(t) e^{\boldsymbol{x}_{2i}(t)'\boldsymbol{\beta_2} + \gamma W_{1i}(t)} dA_0(t) \right\}$$

and we note for future use that:

$$\frac{\partial L_\gamma}{\partial \gamma} = U_\gamma(\tau) L_\gamma$$

The derivative of the survival part of the full joint log likelihood function is expressed as follows:

$$\frac{\partial \ell_\gamma}{\partial \gamma} = \frac{\partial \log L_\gamma}{\partial \gamma} = \frac{U_\gamma(\tau) L_\gamma}{L_\gamma} = U_\gamma(\tau)$$

The resulting score test statistic for $\gamma = 0$ is:

$$\begin{aligned} U(\tau) &= E_{W_1|Y}[U_0(\tau)] \\ &= E_{W_1|Y}\left[\sum_{i=1}^{n}\left\{\int_0^\tau W_{1i}(t)d\boldsymbol{N_i}(t) - \int_0^\tau W_{1i}(t)\boldsymbol{H_i}(t)e^{\boldsymbol{x}_{2i}(t)'\boldsymbol{\beta_2}}dA_0(t)\right\}\right] \\ &= \sum_{i=1}^{n}\int_0^\tau E_{W_1|Y}\left[W_{1i}(t)\right]d\boldsymbol{M_i}(t) \end{aligned} \tag{4.6}$$

where

$$\boldsymbol{M_i}(t) = \boldsymbol{N_i}(t) - \boldsymbol{\Lambda_i}(t) = \boldsymbol{N_i}(t) - \int_0^t \boldsymbol{H_i}(u)e^{\boldsymbol{x}_{2i}(u)'\boldsymbol{\beta_2}}dA_0(u)$$

is the usual counting process martingale for the i^{th} individual, and the expectation is with respect to the conditional distribution of the random effects given the longitudinal measurements. Now we consider $U(\tau)$ to be a particular value of a process $\{U(s) : s > 0\}$, and consider W_1 to be predictable.

Under these assumptions, the variance of $U(s)$ can be estimated by:

$$V_1 = \sum_{i=1}^{n}\int_0^\tau E_{W_1|Y}\left[W_{1i}(t)\right]^2 d\boldsymbol{\Lambda_i}(t) \tag{4.7}$$

If the individuals are independent, and under mild conditions, the martingale central limit theorem follows standard normal distribution asymptotically, with the implication that $U(\tau)/\sqrt{V_1(\tau)}$ under H_0 as $n \to \infty$ as the general score statistic. Alternatively:

$$V_2 = \sum_{i=1}^{n}\left\{\int_0^\tau E_{\boldsymbol{W_1}|Y}\left[\boldsymbol{W_{1i}^2}(\boldsymbol{t})\right]d\boldsymbol{\Lambda_i}(\boldsymbol{t}) - \int_0^\tau\int_0^\tau Cov_{\boldsymbol{W_1}|Y}(\boldsymbol{W_{1i}}(\boldsymbol{t}), \boldsymbol{W_{1i}}(\boldsymbol{s}))d\boldsymbol{M_i}(\boldsymbol{t})d\boldsymbol{M_i}(\boldsymbol{s})\right\}$$

if the individuals are independent and with the limiting distribution of $U(\tau)/\sqrt{V_2(\tau)}$ being $N(0,1)$ under mild conditions (Andersen et al., 1993). However, the simulations will be conducted based only on V_1 throughout this thesis, since Henderson et al. (2002) showed that V_1 and V_2 perform similarly, and the authors recommended the simpler one.

4.4 Simulation studies

In order to demonstrate the properties of the proposed score test for the association between the longitudinal measurements and survival times, several

simulation studies were carried out with four types of latent processes, namely Models A, B, C and D, introduced earlier. A comparison is also made between the power of the test based on the theoretical variance and bootstrap variance estimators. The point is that the martingale variance estimators might not hold for the small sample sized studies, so the bootstrap variance estimator represents an alternative way and may be more robust to small samples. However, it is more computationally intensive.

For the simulation studies, a single group of subjects with discrete measurement times was considered. The longitudinal measurements were generated according to (4.1), in which there was one continuous covariate (x_1) and one binary indicator (x_2), along with the indicator of observation times for the fixed covariates. The `simjoint()` function in the `joineR` package in the `R` program was used to generate our hypothetical example (Philipson et al., 2017). This function simulates data from a standard joint model, in a similar manner to the set-up considered in Henderson et al. (2000). It first simulates the longitudinal data for all possible follow-up times using random draws for the multivariate Gaussian random effects and residual error terms.

Here, the model for latent association is assumed to be either random intercept only, or random intercept and slope. The order of the $\boldsymbol{\beta_1}$ vector in each row is an intercept, a continuous covariate, a binary covariate, and the measurement time (i.e., $\boldsymbol{\beta_1} = (1, x_1, x_2, t)$). The continuous covariate has standard normal distribution, the binary covariate has Binomial distribution with 0.5 probability (as stated in Chapter 3), and the truncation time is chosen 8. However, as stated in the previous chapter `simjoint()` function only simulates data from a subject-level random effects. As such, this function was adapted to be able to simulate data from subject and centre-levels random effects, as in Chapter 3. The key difference between the original function and the modified version is that the modified function simulates multi-level longitudinal data for all possible follow-up times using random draws for the multivariate Gaussian random effects for the centre-level effects.

The survival times were generated according to (4.2), which is an exponential distribution of the baseline hazard. The continuous variable (x_1) and the binary indicator (x_2) were considered once again for each subject with corresponding regression coefficient vector $\boldsymbol{\beta_2}$. The simulation method of the censoring times is the same as that applied in Section 3.5. An exponential distribution of the baseline hazard was specified for Models A and C, and Gompertz distribution of the baseline hazard for Models B and D, as stated in the simulation section of the previous chapter.

Three different scenarios were considered in the simulation of the failure times: a 70% event rate, a 20% event rate with a relatively small subject variability, and 20% event rate and a high subject variability in order to mimic the SLS data. These are referred to here as Scenarios I, II and III, respectively. Different event rates were achieved by choosing specific hazard rates and some other parameters to adapt the scenarios given in the tables under the relevant plots. The chosen parameters are the variance-covariance matrix of random effects, the residual standard errors, and both θ_0 and θ_1.

Each model of each scenario includes 100 simulations, with 201 different γ values ranging from -1 to 1, by an increment 0.01. All these simulations were computationally intensive due to the calculation of bootstrap variance for each model fit. This is especially true when the model becomes time-dependent in the latent process, as the simulations subsequently grow increasingly burdensome. These simulations were performed by remotely connecting to the Northumbria University's Department of Mathematics, Physics and Electrical Engineering server,named Cauchy. The function which took the longest computation time was the bootstrap function, the performance of which we sought to optimise. To do so, the `mclapply()` function in the `parallel` package was used (R Core Team, 2017). Using multiple cores (24), the simulations were performed up to 23 times faster than a computer which has only two cores. Despite this, the simulations took a considerable time. For instance, the simulations of

Model A and Model C (time-independent) were completed in approximately 20 hours, while the completion of the simulations of Model B and Model D took around 90 hours.

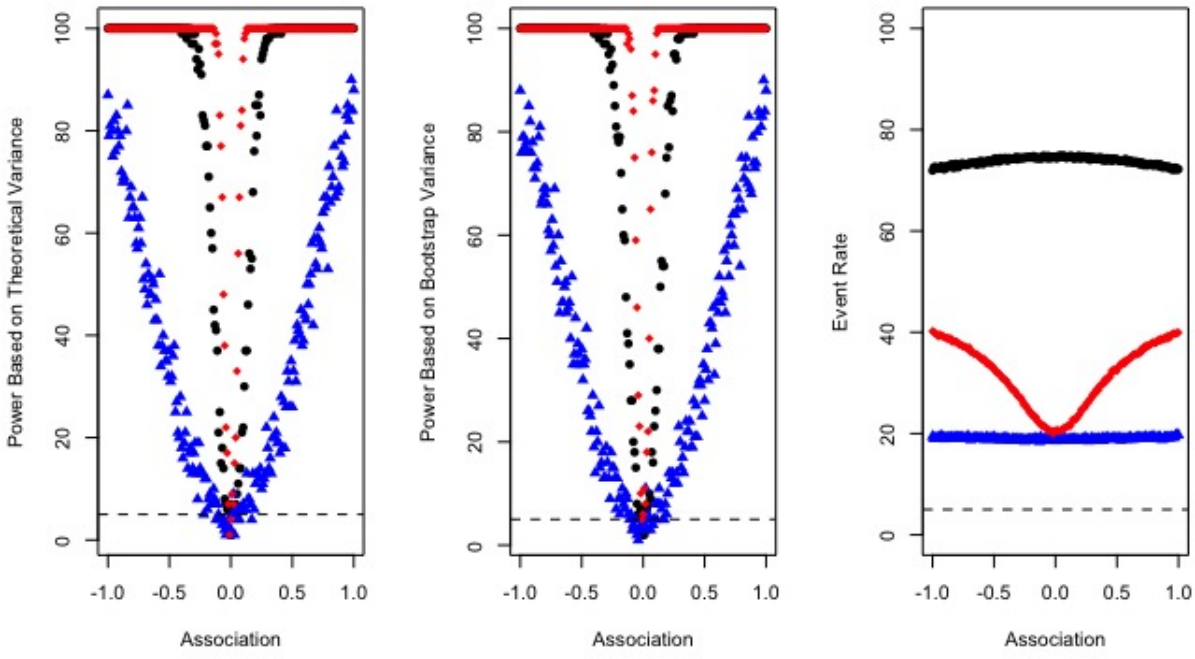

Figure 4.1: Power and event rate for Model A: ● (black) represent a high event rate (Scenario I); ▲ (blue) a low event rate (Scenario II); and ♦(red) mimicked SLS data (Scenario III). Theoretical variance (left panel), bootstrap variance (mid panel), event rate (right panel) are also shown.

Figures 4.1, 4.2, 4.3 and 4.4 provide the estimated power for a nominal 5% test of $H_0 : \gamma = 0$ for the true values of γ, depending on the latent association structures. The power, the proportion of rejecting the null hypothesis, in each case is based on 100 simulations with $n = 500$ subjects. In case of Models C and D, the number of centres was set to 20 ($H = 20$), and the number of subjects at each centre was set to 25 ($m = 25$), resulting in 20 random effects at the centre level, with each repeated 25 times to simplify them, and be able to put them in a single-matrix form, as such, this resulted in 500 random effects at both subject and centre levels for Models C and D. A single γ parameter was assumed for all models, but the approach outlined here can be generalised. All simulations were based on 100 bootstrap samples. As in Henderson et al.

	Model A		
Variable	Scenario-I	Scenario-II	Scenario-III
$\boldsymbol{\beta_1'}$	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)
$\boldsymbol{\beta_2'}$	(0.6, -0.6)	(0.6, -0.6)	(0.6, -0.6)
σ_{u0}	0.707	0.316	4.472
σ_ε	0.1	0.1	5.8
θ_0	-1	-3.3	-3.2

Table 4.1: Selection of the chosen parameters to achieve the relevant event rate, and to adapt the simulations to the scenarios for Model A.

	γ	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)
Scenario I	-1.00	100	100	71.93
	-0.50	100	100	73.69
	0	7	7	74.64
	0.50	100	100	74.10
	1.00	100	100	72.22
Scenario II	-1.00	87	88	19.45
	-0.50	37	38	19.37
	0	2	3	18.57
	0.50	26	28	18.96
	1.00	88	88	19.68
Scenario III	-1.00	100	100	40.22
	-0.50	100	100	32.76
	0	4	6	20.37
	0.50	100	100	32.39
	1.00	100	100	39.92

Table 4.2: The power of the score test expressed as a percentage for Model A, with a representative small sample of the simulation results for each scenario. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

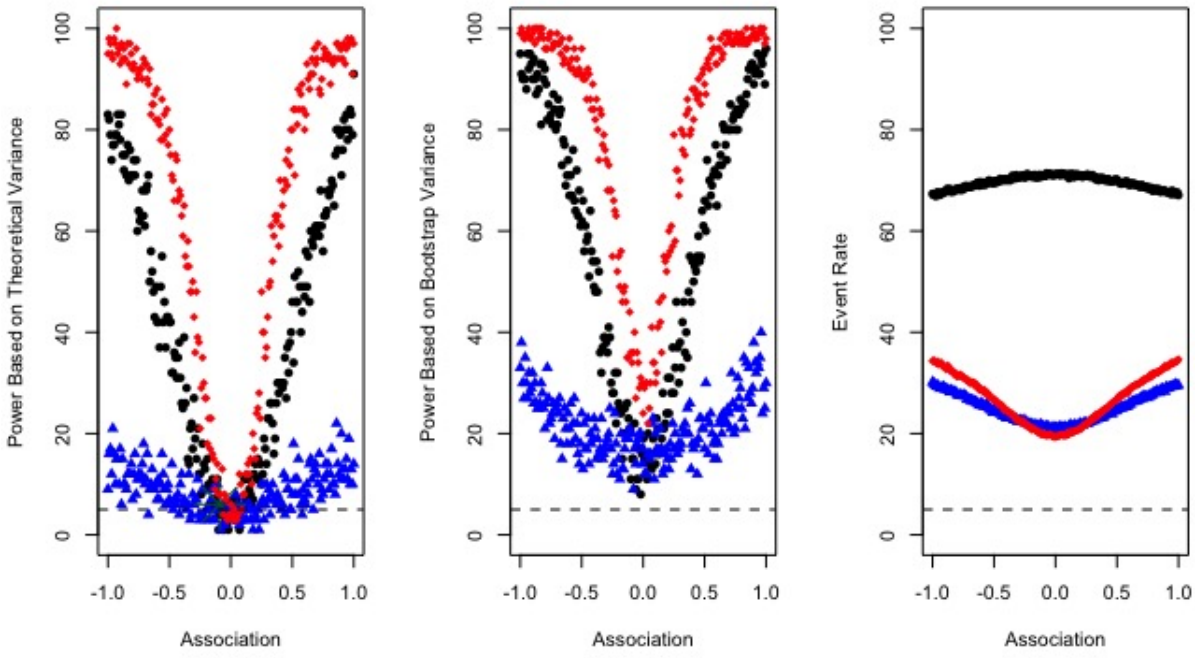

Figure 4.2: Power and event rate for Model B: • (black) represents a high event rate (Scenario I); ▲ (blue) a low event rate (Scenario II); and ♦(red) mimicked SLS data (Scenario III). Theoretical variance (left panel), bootstrap variance (mid panel), and event rate (right panel) are also shown.

	Model B		
Variable	Scenario-I	Scenario-II	Scenario-III
β_1'	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)
β_2'	(0.6, -0.6)	(0.6, -0.6)	(0.6, -0.6)
σ_{u0}^2	0.09	0.1	2.96
σ_{u01}	0.02	0.02	-0.03
σ_{u1}^2	0.04	0.1	0.193
σ_ε	0.1	0.1	4.89
θ_0	-1.9	-4	-4.1
θ_1	0.2	0.2	0.2

Table 4.3: Selection of the chosen parameters to achieve the relevant event rate, and to adapt the simulations to the scenarios for Model B.

	γ	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)
Scenario I	-1.00	83	95	67.34
	-0.50	42	62	69.44
	0	6	19	71.14
	0.50	46	66	70.13
	1.00	91	96	67.08
Scenario II	-1.00	9	33	30.17
	-0.50	7	24	24.88
	0	4	17	21.18
	0.50	8	17	24.60
	1.00	14	30	29.59
Scenario III	-1.00	95	99	34.33
	-0.50	80	96	26.85
	0	6	24	19.35
	0.50	80	89	26.68
	1.00	97	98	34.65

Table 4.4: The power of the score test expressed as a percentage for Model B, with a representative small sample of the simulation results for each scenario. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

	Model C		
Variable	Scenario-I	Scenario-II	Scenario-III
$\boldsymbol{\beta_1'}$	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)
$\boldsymbol{\beta_2'}$	(0.6, -0.6)	(0.6, -0.6)	(0.6, -0.6)
σ_{u0}	0.1	0.1	4.7
σ_{v0}	0.1	0.1	0.12
σ_{ε}	0.1	0.1	5.88
θ_0	-1	-3.2	-3.2

Table 4.5: Selection of the chosen parameters to achieve the relevant event rate, and to adapt the simulations to the scenarios for Model C.

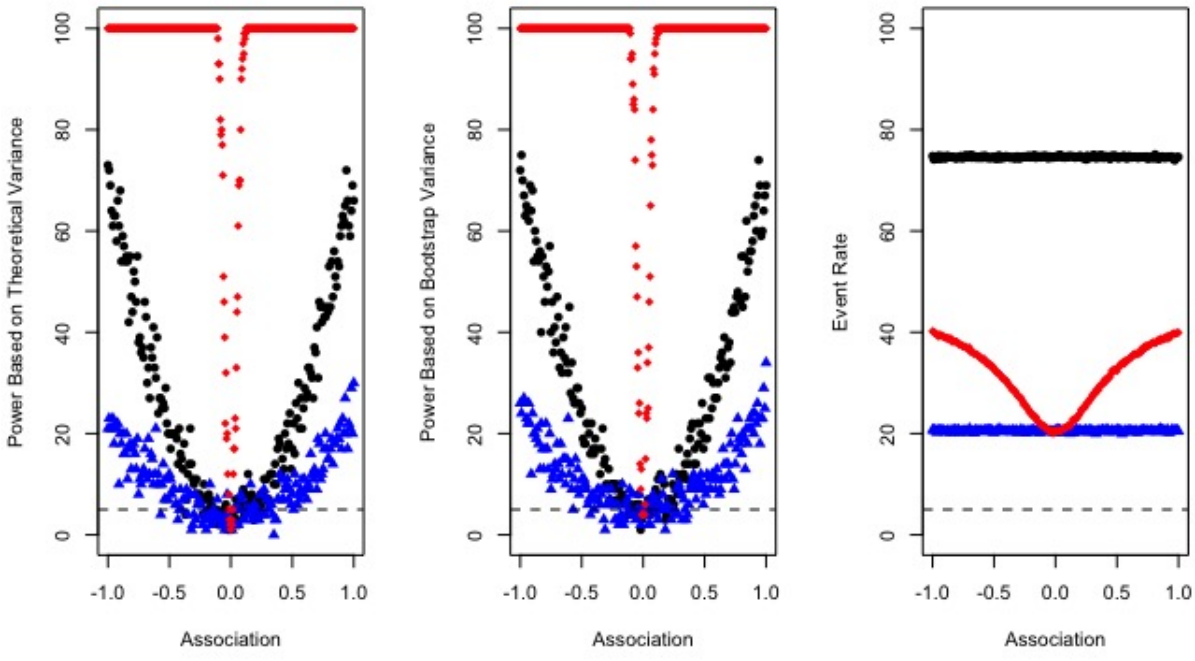

Figure 4.3: Power and event rate for Model C: ● (black) represent a high event rate (Scenario I); ▲ (blue) a low event rate (Scenario II); and ♦(red) mimicked SLS data (Scenario III). Theoretical variance (left panel), bootstrap variance (mid panel), event rate (right panel) are also shown.

	γ	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)
Scenario I	-1.00	73	72	74.80
	-0.50	20	25	74.62
	0	1	5	74.71
	0.50	17	19	74.60
	1.00	66	69	74.62
Scenario II	-1.00	21	26	20.71
	-0.50	12	11	20.18
	0	3	4	20.68
	0.50	8	29	20.78
	1.00	30	34	20.43
Scenario III	-1.00	100	100	39.66
	-0.50	100	100	32.42
	0	1	4	20.43
	0.50	100	100	33.11
	1.00	100	100	39.73

Table 4.6: The power of the score test expressed as a percentage for Model C, with a representative small sample of the simulation results for each scenario. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

	Model D		
Variable	Scenario-I	Scenario-II	Scenario-III
$\boldsymbol{\beta_1'}$	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)	(7, -1, 0.5, 0.1)
$\boldsymbol{\beta_2'}$	(0.6, -0.6)	(0.6, -0.6)	(0.6, -0.6)
σ^2_{u0}	0.3	0.3	18
σ_{u01}	0.1	0.1	-0.03
σ^2_{u1}	0.2	0.2	0.193
σ^2_{v0}	0.2	0.2	2
σ_{v01}	0.001	0.001	0.001
σ^2_{v1}	0.1	0.1	0.1
σ_ε	0.1	0.1	5.00
θ_0	-1.9	-5.5	-5.5
θ_1	0.2	0.5	0.5

Table 4.7: Selection of the chosen parameters to achieve the relevant event rate, and to adapt the simulations to the scenarios for Model D.

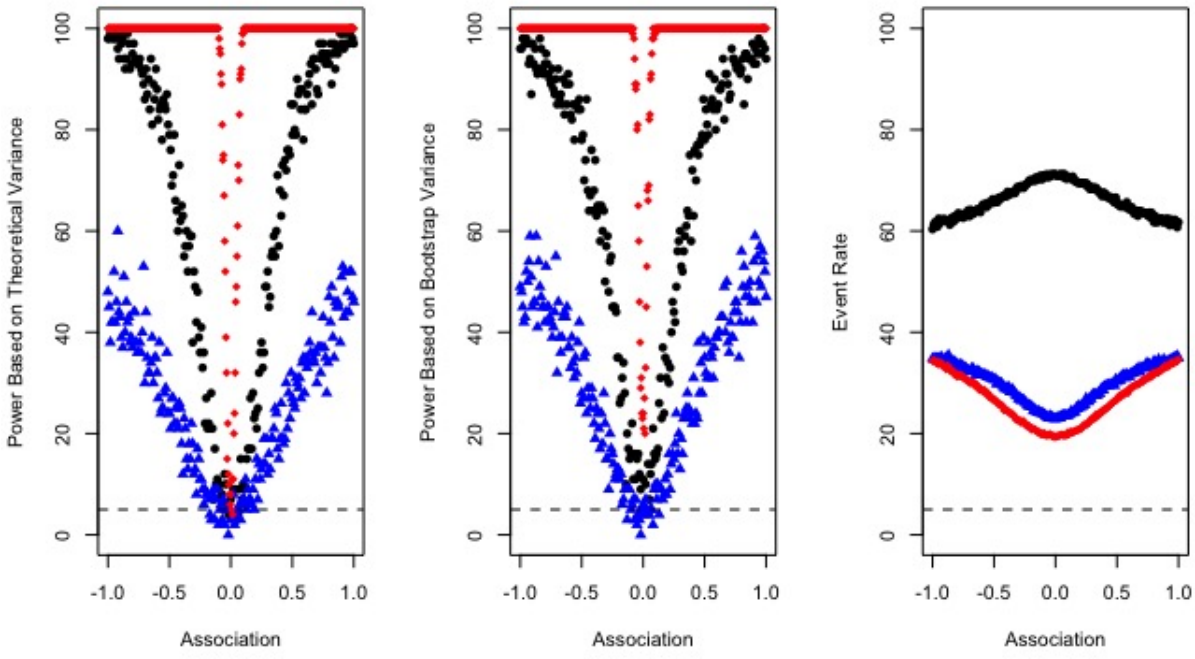

Figure 4.4: Power and event rate for Model D: ● (black) represents a high event rate (Scenario I); ▲ (blue) a low event rate (Scenario II); ♦(red) mimicked the SLS data (Scenario III). Theoretical variance (left panel), bootstrap variance (mid panel), event rate (right panel) are also shown.

	γ	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)
Scenario I	-1.00	98	96	60.36
	-0.50	79	78	65.62
	0	7	11	71.18
	0.50	79	78	65.53
	1.00	97	94	61.74
Scenario II	-1.00	48	49	38.82
	-0.50	24	28	30.62
	0	3	6	22.89
	0.50	27	35	30.25
	1.00	46	47	34.89
Scenario III	-1.00	100	100	33.25
	-0.50	100	100	27.12
	0	8	24	23.10
	0.50	100	100	26.02
	1.00	100	100	34.82

Table 4.8: The power of the score test expressed as a percentage for Model D, with a representative small sample of the simulation results for each scenario. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

(2002), the parameter values required for the test statistic were replaced with the estimates of the separate analysis of longitudinal measurements and time-to-event. These were analysed using the `lme()` function in the `nlme` package the Pinheiro et al. (2017), and the `coxph()` function in the `survival` package Therneau (2015) in R. Tables 4.1, 4.3, 4.5 and 4.7 represent the specifically chosen parameters for each model. On the other hand, Tables 4.2, 4.4, 4.6 and 4.8 depict the calculations of two different types of power and event rate for a selection of association parameters, typically $\gamma = (-1, -0.5, 0, 0.5, 1)$. Even though each figure represents simulations of 201 γ values from -1 to 1 by an increment of 0.01, each table only shows five of the simulation results with a representative selection of γ parameters. Results are given for the martingale-based variance estimator, V_1, and the bootstrap variance estimator. As stated in Henderson et al. (2002), the martingale-based variance estimator, V_1, and a more complicated variance estimator with correction term V_2 perform similarly, so the simpler V_1 is recommended. Hence, all of the following calculations of the

test statistic is based on V_1 and the bootstrap variance estimator.

Overall, those figures indicate that the score test is less powerful and noisier with a low subject variability and a low event rate than with a low subject variability and a high event rate, and a high subject variability and a low event rate. To be able to capture the right information from the score test, it is expected that at least either the variations of random effects or the event rate is reasonably high. Moreover, when the variation of the random effects is reasonably high, the event rate increases as association gets far away from zero, as seen from the data in the aforementioned figures' right panels, represented by ♦ (red).

From Figure 4.2, the bootstrap variance estimator seems to exhibit some bias. One would expect a few of the points to be under the dashed line due to the variation. However, in this figure, the mid panel shows that all the results are above the dashed line. Bootstrap is therefore systematically underestimating the variance, and the power based on the bootstrap variance is found to be artificially high.

Figures 4.1, 4.2, 4.3 and 4.4 show that the power of the test increases as γ moves away from zero, as one expects for fixed γ values. Power is also associated with the event rate, and highly affected by the variance covariance matrix of the corresponding random effect terms. For example, when Scenario II and Scenario III are compared with relatively close event rate for all models, the score test is found to be more powerful in a narrower γ range. Hence the power of the test in Scenario III is always higher than the power in Scenario II for both variance estimators. To follow, the small variability among the random effects terms and then would like to see the behaviour of the test when we mimic the SLS data. We first deal with the small variability among the random effects terms and then would like to see the behaviour of the test when we mimic the SLS data.

4.4.1 Comparison of power of the score test based on two variance structures

This simulation study aims to investigate how the performances of the theoretical variance and bootstrap variance differ. Figure 4.5, 4.6 and 4.7 provide an overview of the differences between the power of the score test based on the theoretical variance and bootstrap variance. The plot is of, $Power_{V_1} - Power_{boot}$.

What stands out in those figures is that the difference randomly oscillates for Model A and Model B, respectively. However, when the model gets more complicated, the difference first starts to systematically decrease, and then systematically increase, as the association gets closer to zero for Scenario I. The power difference plots of the Scenario II look like random scatter for Models A and B. However, the power is systematically higher for the bootstrap for Models C and D. As mentioned previously, the bootstrap variance estimator underestimates the variance, and is deriving the power to be artificially large.

4.5 Application: the scleroderma lung study

The application in this chapter, once more, utilises the scleroderma lung study introduced in Section 1.5.1. A joint model of multi-level repeated measurements and survival outcomes was fitted to this data in the previous chapter. The score test developed in this chapter is now applied to this dataset.

The longitudinal and the survival data was first analysed separately using the `R` packages `nlme` and `survival`, and their respective functions therein `lme()` and `coxph()`, respectively.

The standardized score statistics, $U/V^{1/2}$, using the martingale variance formulation and bootstrap variance are shown in Table 4.9. The table indicates that only Model A has very strong evidence of association between survival time and FVC index, as shown by the p-values. The rest of the models have no

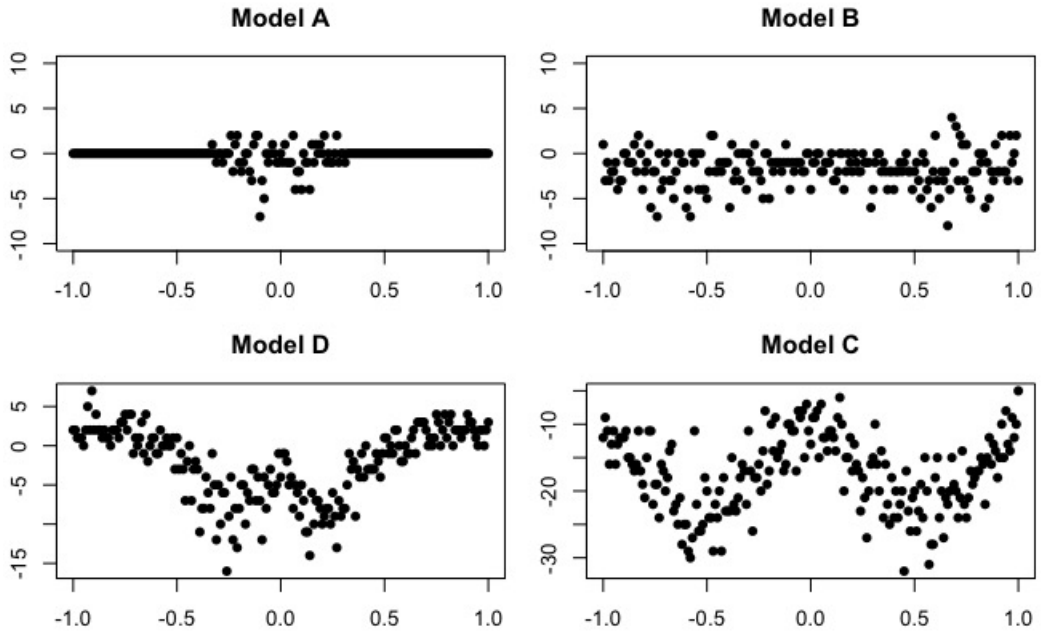

Figure 4.5: Difference in the score test power with various latent associations (on the x-axis) for Scenario I and Models A-D (clockwise from top-left).

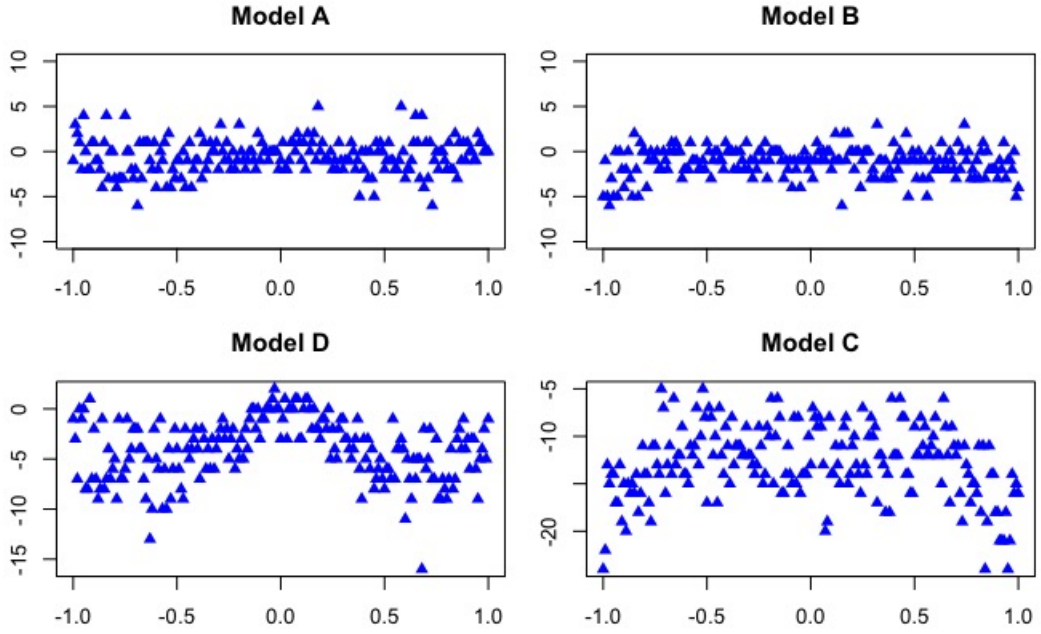

Figure 4.6: Difference in the score test power with various latent associations (on the x-axis) for Scenario II and Models A-D (clockwise from top-left).

	Score statistics (martingale variance)	p-value	Score statistics (bootstrap variance)	p-value
Model A	-2.925*	0.003*	-2.400*	0.016*
Model B	0.108	0.914	0.094	0.925
Model C	-0.872	0.383	-1.115	0.265
Model D	-1.136	0.256	-1.002	0.316

Table 4.9: Score statistics and p-values of the SLS data using martingale and bootstrap variance

evidence of association between survival and longitudinal outcomes. The score statistics were compared to $N(0,1)$ to obtain an exact p-value.

In addition to the data above, one would like to see how powerful the score test is when a real dataset is used. To do so, the models of the SLS data are first fitted using the method described by Henderson et al. (2000) and the `joineR` package (Philipson et al., 2017), with standard errors attained by bootstrap methods based on re-estimation from simulated data. Table 4.10 presents the parameter estimates, standard errors, and maximized log likelihoods obtained from the joint analyses. Negative estimates of γ confirm that high FVC levels are associated with reduced risk. After fitting a candidate joint model and estimating the association between FVC level and survival, one can investigate how reliable the test may be. Simulations were performed with parameter values chosen to mimic those four under-fitted models, referred to as Scenario III. When $\gamma = -0.088$ and the latent process is the random intercept, the power based on martingale variance is 86, the power based on bootstrap variance is 93, and the mean of the event rate of 100 simulations is 21.38%, which similar to the observed 19.1% in the data. On the other hand, when $\gamma = -0.097$ and the latent process is random intercept and slope, the power based on martingale variance is 5, the power based on bootstrap variance is 37, and the event rate is 20.04%. As such, adding a random slope term to the latent process makes the model more complicated and the score test less powerful. However, the small sample and low event rate may have contributed to this.

	Model A		Model B	
	$W_1(t) = U_1$, $W_2(t) = \gamma U_1$		$W_1(t) = U_1 + U_2 t$, $W_2(t) = \gamma(U_1 + U_2 t)$	
	Est	SE	Est	SE
Longitudinal				
Constant	-4.855	5.070	0.796	2.754
MAXFIB	0.371	0.743	-0.151	0.495
CYC	13.010	7.771	5.802	4.028
months	0.031	0.062	-0.033	0.060
FVC0	1.040	0.066	0.977	0.035
MAXFIB:CYC	-1.722	0.942	-1.004	0.606
CYC:months	-0.139	0.092	-0.112	0.079
CYC:FVC0	-0.131	0.104	-0.048	0.058
Survival				
MAXFIB	0.320	0.211	0.321	0.214
CYC	-0.061	0.370	-0.103	0.369
Association				
γ_1	-0.088	0.083	-0.097	0.093
Random effects				
σ^2_{u1}	20.337	3.267	2.964	1.585
σ^2_{u2}			0.193	0.042
Noise				
σ^2_ε	34.750	5.569	23.940	4.228
Log likelihoods				
Longitudinal	-3502.201		-3366.932	
Survival	-117.527		-116.886	
Combined	-3619.728		-3483.819	

Table 4.10: The SLS data results for two different $W_1(t)$ models: a random intercept model at subject-level (Model A), and a random intercept and slope model at subject-level with proportional association (Model B)

4.6 Discussion

In this chapter, the aim was to concentrate on individual and centre-level surrogacy, in the form of longitudinal repeated measurements, which is valid in the case of strong association between subject values and a survival endpoint (Buyse et al., 2000). There is no current software to account for multilevel joint models, and a score test is an ideal prognostic tool to investigate the necessity of these associations. This brings about the need of a score test for association between longitudinal and survival outcomes to decide whether a joint model is necessary for the analysis. As joint modelling can be a computationally intensive analysis, it is of importance to know which biomarker or biomarkers are needed to include in the model in the case of having multiple longitudinal profiles, the nature of the association structure these biomarkers should have, and, moreover, if a biomarker(s) is actually needed at all.

Simulation studies were presented with three different scenarios and four types of random effects models: Scenario I - low subject variability with $\sim 70\%$ event rate, Scenario II - low subject variability with $\sim 20\%$ event rate, and Scenario III - high subject variability with $\sim 20\%$ event rate. The results demonstrate that as the association gets close to zero, the power of the test decreases in all scenarios, as expected. However the range is narrower when the between-subject variability is high. An analysis of the motivating SLS data was also reported, where the main outcomes were the potential biomarker FVC and all-cause mortality.

One would expect some power performed below the nominal 5% threshold for all models due to variation. However, it was surprising that the power based on the bootstrap variance (Figure 4.2) performed above the nominal 5% threshold, even under the null of no association. In addition, it was difficult to maintain the event rate at a certain level when random slope terms were added in the latent process model.

The same patterns were approximately seen for the differences of the power of the test based on two different variance terms for all the simulation scenarios. When the random slope terms are added, the difference systematically decreases and increases after a certain point as the association approaches zero.

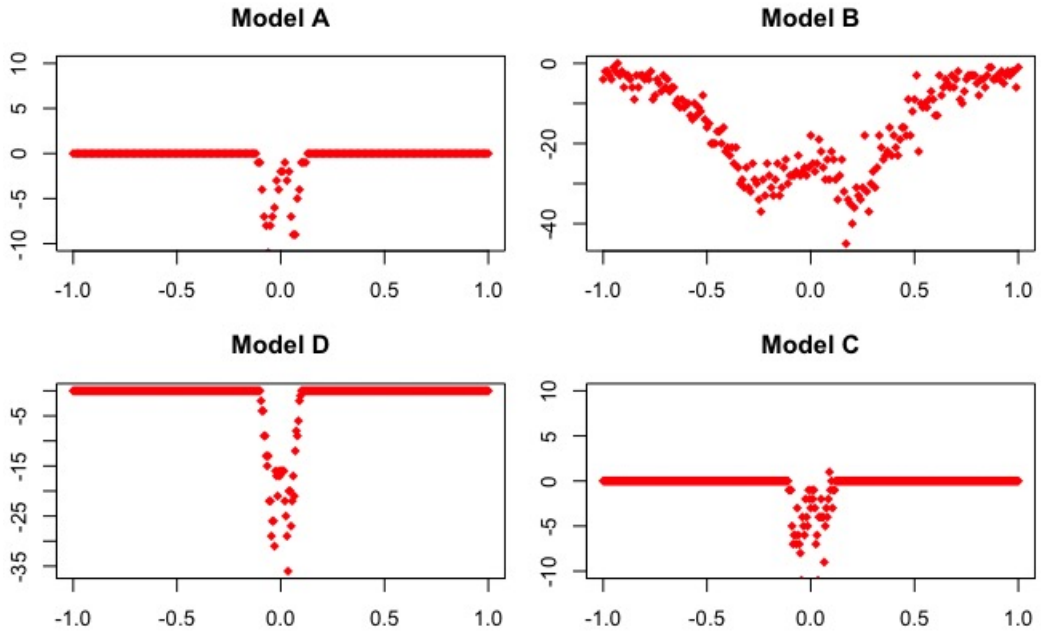

Figure 4.7: Difference in the score test power with various latent associations (on the x-axis) for Scenario III and Models A-D (clockwise from top-left).

Chapter 5

A score test for complex joint models

5.1 Introduction

In the previous chapter, the main concepts of a score test for four different joint models in conjunction with a proportional association (thereby involving a single association parameter) were introduced, simulation studies were conducted and a motivation for the score test approach was described. In this chapter, methodological extensions of the score test for association between longitudinal and survival outcomes were proposed to address the issue of complex association structure between those outcomes. A score test for multiple parameters (i.e., separate association for each random effect term between repeated measurement and event times data), followed by a score test for multivariate joint models and, finally, a score test for multilevel joint models were considered.

There has been little attention on the identification of the longitudinal biomarkers for survival by a score test and they were introduced in the previous chapter. So far this method has only been applied to a single association parameter. In this chapter, firstly, the score test for multiple association parameters derived from joint models with three different latent association structures is extended. Secondly, the ideas to multivariate longitudinal case will be extended to deal with which longitudinal biomarker should be in the model in the case of hav-

ing more than one correlated longitudinal biomarkers. Thirdly, the need for the centre-level effect in multilevel joint models will be investigated.

Several simulation studies with three scenarios will be conducted to assess the performance and power of the various tests and to make comparison between two candidate estimators of the score test statistic: with martingale variance and bootstrap variance as performed in the previous chapter for the first extension. (The performance of the power of the test with only one scenario and based on the Martingale variance estimator for multivariate score test, and, with, again, one scenario based on two variance estimators for multilevel score test were investigated due to computational burden.)

The rest of the chapter is organised as follows: Section 5.2 describes the models and notation used throughout this chapter. The subsequent section gives details on the score test for association in terms of three models that we are interested in. Simulation studies regarding the aforementioned models are presented and the results are discussed in Section 5.4. The application of the SLS data for the identification of the separate effect association is given in Section 5.5. The chapter concludes with a discussion of the extended methods and simulation studies.

5.2 Models and notation

This section explains the models and notation that we have adopted throughout this chapter for all three extensions of the investigation of the score test.

We begin with the model and notation introduced in Section 4.2. We will firstly deal with three different random effect structures, namely Models B, Equation (4.4), C, Equation (3.3) and D, Equation (3.4), random intercept and slope model at subject level (B), random intercept model at both subject and centre levels (C) and random intercept and slope model at both subject and centre levels (D), respectively, as introduced in the previous chapter, for the separate

latent association structure of the random effects.

A multivariate joint model is another extension of the model proposed by Henderson et al. (2000), taking account of multivariate longitudinal data. For this model we introduce a new parameter K as the index of the longitudinal biomarker and $Y_i = (Y_{i1}', \ldots, Y_{iK}')$ is the K-variate continuous outcome vector where Y_{ik} represents an $(n_{ik} \times 1)$-vector of observed repeated outcomes for the k^{th} biomarker i.e., $Y_{ik} = (Y_{i1k}, \ldots, Y_{in_{ik}k})'$. The measurements of these outcomes are taken at t_{ijk} times for $j = 1, \ldots, n_{ik}$ allowing unbalanced data for complete flexibility. The event time for each subject, T_i*, potential censoring time, C_i and the failure indicator, δ_i are as described earlier since we still have univariate survival data.

The zero-mean, $(K+1)$-variate Gaussian latent process, which links the separate submodels, for each subject is $\boldsymbol{W}_i(t) = \left\{\boldsymbol{W}_{1i}^{(1)}(t), \ldots, \boldsymbol{W}_{1i}^{(K)}(t), \boldsymbol{W}_{2i}(t)\right\}$.

The longitudinal submodels for the multivariate biomarkers are given by

$$Y_{ik}(t) = \boldsymbol{X}_{1ik}'(t)\boldsymbol{\beta_{1k}} + \boldsymbol{W}_{1i}^{(k)}(t) + \varepsilon_{ik} \tag{5.1}$$

where $\boldsymbol{X}_{1ik}(t)$ is a p_k-vector of potentially time varying covariates, $\boldsymbol{\beta}_{1k}$ is the corresponding regression coefficients and $\boldsymbol{W}_{1ik}^{(k)}(t)$ is given by

$$\boldsymbol{W}_{1i}^{(k)}(t) = \boldsymbol{D}_{ik}'(t)\boldsymbol{U}_{ik} \quad \boldsymbol{U}_{ik} \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma_{kk}}) \tag{5.2}$$

where $\boldsymbol{D}_{ik}(t)$ is an r_k-vector of potentially time varying covariates, $\boldsymbol{U_{ik}}$ are random effect terms, Σ_{kk} is an $(r_k \times r_k)$-variance-covariance matrix. We also consider the dependency between biomarkers, let $Cov(\boldsymbol{U}_{ik}, \boldsymbol{U}_{il}) = \Sigma_{kl}$ for $k \neq l$. Additionally, ε_{ik} and $\boldsymbol{U}_{ik}$ are assumed to be uncorrelated as a standard modelling assumption. The assumptions made here for the models (5.1) and (5.2) are equivalent to the multivariate extension of the Laird-Ware model (Laird and Ware, 1982).

The time-to-event submodel is the same (Equation (4.2)) as that described earlier in Section 4.2. Here, the $\boldsymbol{W_{2i}(t)}$ term is a linear combination of $\left\{\boldsymbol{W}_{1i}^{(1)}(t), \ldots, \boldsymbol{W}_{1i}^{(K)}(t)\right\}$

with the vector of association parameters $\boldsymbol{\gamma}_Y = (\gamma_{Y1}, \ldots, \gamma_{YK})$. The estimation method is the maximum likehood estimation method using the EM algorithm is explained in detail by Hickey et al. (2018).

Thirdly, our aim is to investigate if we need higher cluster levels of random effects. To do so, our interest will be on the Models C and D with proportional association within levels in the former and, separate association within levels in the latter at both level. We use the longitudinal submodel and notation introduced in Section 3.2 with Equation (3.1) and the survival submodel with intensity introduced in Section 4.2, equation (4.2). Here the latent association structures in the survival submodel is

$$\text{Model C:} \quad \boldsymbol{W_{2hi}(t)} = \gamma_1(u_{0i}) + \gamma_2(v_{0h}) \tag{5.3}$$

$$\text{Model D:} \quad \boldsymbol{W_{2hi}(t)} = \gamma_1(u_{0i} + u_{1i} * t) + \gamma_2(v_{0h} + v_{1h} * t) \tag{5.4}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ is the vector of latent association parameters and the random effect models in the longitudinal submodel are specified as (3.3) and (3.4) for Models C and D, respectively.

5.3 Score test for association

The score tests introduced here are the aforementioned extension of the score test introduced in Section 4.3.

5.3.1 Score test for the separate association parameter

We firstly extend the score test under the null hypothesis $H_0 : \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q) = \mathbf{0}$ where q is the number of random effects and H_1: At least one component of $\boldsymbol{\gamma}$ parameter is different from zero. The resulting score test statistic under the null hypothesis is calculated as Equation (4.6) and the Martingale variance is calculated with Equation (4.7). The key difference is the score statistic for multiple parameters which is calculated as:

$$S_{sep} = \boldsymbol{U}(\boldsymbol{\tau})' \boldsymbol{V_1}(\boldsymbol{\tau})^{-1} \boldsymbol{U}(\boldsymbol{\tau}) \tag{5.5}$$

where τ is the maximum follow-up time and this test statistic follows χ^2_m where m is the number of constraints imposed by the null hypothesis. Separate analysis of each component will again be used to calculate the score statistic.

5.3.2 Score test for association of multivariate joint model

This section is an extension of the score test for association proposed by Henderson et al. (2002), also explained in detail in the previous chapter, to the case of joint model with multivariate longitudinal measurements and a single event times data. Now, this test is no longer based on separate analysis of the two components since now, under the null, we still have a joint model. The test is still based on a multivariate joint model without having the account of k^{th} longitudinal biomarker under the null hypothesis $H_0 : \gamma_{Yk} = 0$, for $1 < k < K$, where K is the number of longitudinal biomarkers.

Let the collection of unknown parameters be $(\boldsymbol{\theta}, \boldsymbol{\gamma_Y}, \boldsymbol{\beta_2}, A_0)$, where $\boldsymbol{\theta}$ includes all parameters estimated with the distribution of Y_k and $\boldsymbol{\gamma}_Y = (\gamma_{Y1}, \ldots, \gamma_{YK})$ as specified earlier. Let the maximum follow-up time be τ. The conditional likelihood of the event history data can be written as

$$L_{\gamma_{Yk}} = \left(\prod_t \prod_i \left(e^{\boldsymbol{X_{2i}(t)'\beta_2} + \sum_{k=1}^{K} \gamma_{Yk} \boldsymbol{W_{1i}^{(k)}(t)}} dA_0(t) \right)^{\Delta \boldsymbol{N_i(t)}} \right) \times \exp\left(- \int_0^\tau \sum_{i=1}^{n_k} \boldsymbol{H_i(t)} e^{\boldsymbol{X_{2i}(t)'\beta_2} + \sum_{k=1}^{K} \gamma_{Yk} \boldsymbol{W_{1i}^{(k)}(t)}} dA_0^{(k)}(t) \right) \quad (5.6)$$

where $\boldsymbol{N_i}$ is the counting process for the i^{th} subject and $A_0^{(k)}$ is the cumulative baseline hazard predicted from the joint model of multivariate longitudinal meaurements of K except k.

$$A_0^{(k)}(u) = \int_0^u \frac{\boldsymbol{J^{(k)}(s)}}{\sum_{i=1}^{n} \boldsymbol{H_i^{(k)}(s)} e^{\boldsymbol{X_{2i}(s)'\beta_2} + \sum_k \gamma_{Y(-k)} \boldsymbol{W_{1i}^{(-k)}(t)}}} d\boldsymbol{N(s)},$$

where $\boldsymbol{N(s)} = \sum \boldsymbol{N_i(s)}$, $\boldsymbol{J^{(k)}(s)} = \boldsymbol{I}\left[\sum \boldsymbol{H_i^{(k)}(s)} > 0\right]$ and $\gamma_{Y(-k)}$ is the association parameters of random effects excluding k between 1 and K under the null hypothesis, and $\boldsymbol{W_{1i}^{(-k)}}$ is the vector of random effects for the i^{th} subject regarding the biomarkers between 1 and K except k^{th} biomarker.

Let

$$U_{\gamma_{Yk}}(\tau) = \sum_{i=1}^{n_k} \left\{ \int_0^\tau \boldsymbol{W_{1i}^{(k)}}(\boldsymbol{t}) d\boldsymbol{N_i}(\boldsymbol{t}) - \int_0^\tau \boldsymbol{W_{1i}^{(k)}}(t) \boldsymbol{H_i^{(k)}}(\boldsymbol{t}) e^{\boldsymbol{X}_{2i}(\boldsymbol{t})'\boldsymbol{\beta_2} + \sum_{k=1}^{K} \gamma_{Yk} \boldsymbol{W_{1i}^{(k)}}(\boldsymbol{t})} dA_0^{(k)}(t) \right\}$$

and we note

$$\frac{\partial L_{\gamma_{Yk}}}{\partial \gamma_{Yk}} = U_{\gamma_{Yk}}(\tau) L_{\gamma_{Yk}}$$

The derivative of the survival part of the full joint log-likelihood function is indicated as

$$\frac{\partial \ell_{\gamma_{Yk}}}{\partial \gamma_{Yk}} = \frac{\partial \log L_{\gamma_{Yk}}}{\partial \gamma_{Yk}} = \frac{U_{\gamma_{Yk}}(\tau) L_{\gamma_{Yk}}}{L_{\gamma_{Yk}}} = U_{\gamma_{Yk}}(\tau)$$

The resulting score test statistic for $\gamma_{Yk} = 0$ for $1 < k < K$, $U^{(k)}(\tau)$ is

$$\begin{aligned}
&U^{(k)}(\tau) \\
&= E_{\boldsymbol{W_1^{(k)}}|Yk}[U_0^{(k)}(\tau)] \\
&= E_{\boldsymbol{W_1^{(k)}}|Yk}\left[\sum_{i=1}^{n_k} \left\{ \int_0^\tau \boldsymbol{W_{1i}^{(k)}}(\boldsymbol{t}) d\boldsymbol{N_i}(\boldsymbol{t}) - \int_0^\tau \boldsymbol{W_{1i}^{(k)}}(t) \boldsymbol{H_i^{(k)}}(\boldsymbol{t}) e^{\boldsymbol{x}_{2i}(\boldsymbol{t})'\boldsymbol{\beta_2} + \sum_k \gamma_{Y(-k)} \boldsymbol{W_{1i}^{(-k)}}(\boldsymbol{t})} dA_0^{(k)}(t) \right\} \right] \\
&= \sum_{i=1}^{n_k} \int_0^\tau E_{\boldsymbol{W_1^{(k)}}|Yk}\left[\boldsymbol{W_{1i}^{(k)}}(\boldsymbol{t}) \right] d\boldsymbol{M_i^{(k)}}(\boldsymbol{t})
\end{aligned}$$

where

$$\boldsymbol{M_i^{(k)}}(\boldsymbol{t}) = \boldsymbol{N_i}(\boldsymbol{t}) - \boldsymbol{\Lambda_i^{(k)}}(\boldsymbol{t}) = \boldsymbol{N_i}(\boldsymbol{t}) - \int_0^t \boldsymbol{H_i^{(k)}}(\boldsymbol{u}) e^{\boldsymbol{x}_{2i}(\boldsymbol{u})'\boldsymbol{\beta_2} + \sum_k \gamma_{Y(-k)} \boldsymbol{W_{1i}^{(-k)}}(\boldsymbol{t})} dA_0^{(k)}(u)$$

is the usual counting process martingale for the i^{th} subject and k^{th} longitudinal biomarker and the expectation is with respect to the conditional distribution of the random effects given the corresponding longitudinal measurements. Now we consider $\boldsymbol{W_1^{(k)}}$ to be predictable and the variance of $U^{(k)}$ can be estimated by

$$V_1^{(k)} = \sum_{i=1}^{n_k} \int_0^\tau E_{\boldsymbol{W_1^{(k)}}|Yk}\left[\boldsymbol{W_{1i}^{(k)}}(\boldsymbol{t}) \right]^2 d\boldsymbol{\Lambda_i^{(k)}}(\boldsymbol{t}) \tag{5.7}$$

Under mild conditions with independency between individuals, the martingale central limit theorem follows asymptotically with the implication that $U^{(k)}(\tau)/\sqrt{V_1^{(k)}(\tau)}$ is a standard normal random variable under H_0 as $n_k \to \infty$(Andersen et al., 1993).

5.3.3 Score test for the cluster level association

Although the score test is based on the separate analysis of the two components (longitudinal measurements and event times), it still remains as a joint model while testing the higher cluster level i.e. without a centre level effect we have a traditional joint model at the individual level. We deal with the latent association in the time-to-event submodels with the proportional association within the same level random effects, separate association across levels, specified as equations (5.3) and (5.4). However in the development below we use only the latter model as it is more comprehensive than the former.

This section is the extension of the score test to consider association in the multilevel joint model proposed earlier in Chapter 3 and also explained in detail in the previous chapter to the case of joint model with multilevel longitudinal measurements and a single event times data. The test is still based on a joint model accounting for the subject level random effects. The null hypothesis is $H_0 : \gamma_2 = 0$.

Let the collection of unknown parameters be $(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta_2}, A_0)$, where $\boldsymbol{\theta}$ includes all parameters estimated with the distribution of $\boldsymbol{Y_h}$ this time. Let the maximum follow-up time be τ. The conditional likelihood of the event history data can be written in a similar way to the conditional likelihood in the previous extension:

$$L_\gamma = \left(\prod_t \prod_i \left(e^{\boldsymbol{x_{2i}(t)'\beta_2 + W_{2hi}(t)}} dA_0(t)\right)^{\boldsymbol{\Delta N_i(t)}} \right) \times \exp\left(- \int_0^\tau \sum_{i=1}^{n} \boldsymbol{H_i(t)} e^{\boldsymbol{X_{2i}(t)'\beta_2 + W_{2hi}(t)}} dA_0(t) \right) \quad (5.8)$$

where $\boldsymbol{W_{2hi}(t)} = \gamma_1(u_{0i} + u_{1i} * t) + \gamma_2(v_{0h} + v_{1h} * t)$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, $\boldsymbol{N_i}$ is the counting process for the i^{th} subject and A_0 is the cumulative baseline hazard predicted from the multilevel joint model.

In practice the unknown parameter $\hat{\beta}_2$ is replaced by its maximum partial likelihood estimator under H_0, and the random effects in time-to-event is only re-

lated to the subject level under H_0.

$$A_0(u) = \int_0^u \frac{\boldsymbol{J}(\boldsymbol{s})}{\sum_{i=1}^n \boldsymbol{H_i}(\boldsymbol{s}) e^{\boldsymbol{x}_{2i}(\boldsymbol{s})'\boldsymbol{\beta_2} + \gamma_1(u_{0i} + u_{1i} * t)}} d\boldsymbol{N}(\boldsymbol{s}),$$

where $\boldsymbol{N}(\boldsymbol{s}) = \sum \boldsymbol{N_i}(\boldsymbol{s})$, $\boldsymbol{J}(\boldsymbol{s}) = \boldsymbol{I}[\sum \boldsymbol{H_i}(\boldsymbol{s}) > 0]$ and γ_1 is the association parameter for subject level random effects.

Let

$$U_\gamma(\tau) = \sum_{i=1} \left\{ \int_0^\tau \boldsymbol{W_{1hi}}(\boldsymbol{t}) d\boldsymbol{N_i}(\boldsymbol{t}) - \int_0^\tau \boldsymbol{W_{1hi}}(t) \boldsymbol{H_i}(\boldsymbol{t}) e^{\boldsymbol{x}_{2i}(\boldsymbol{t})'\boldsymbol{\beta_2} + \boldsymbol{W}_{2hi}(\boldsymbol{t})} dA_0(t) \right\}$$

and we note

$$\frac{\partial L_\gamma}{\partial \gamma} = U_\gamma(\tau) L_\gamma$$

The derivative of the survival part of the full joint log-likelihood function is indicated as

$$\frac{\partial \ell_\gamma}{\partial \gamma} = \frac{\partial \log L_\gamma}{\partial \gamma} = \frac{U_\gamma(\tau) L_\gamma}{L_\gamma} = U_\gamma(\tau)$$

The resulting score test statistic for $\gamma_2 = 0$, $U(\tau)$ is

$$\begin{aligned}
&U(\tau) \\
&= E_{\boldsymbol{W_1}|Y}[U_0(\tau)] \\
&= E_{\boldsymbol{W_1}|Y}\left[\sum_{i=1}^n \left\{ \int_0^\tau \boldsymbol{W_{1hi}}(\boldsymbol{t}) d\boldsymbol{N_i}(\boldsymbol{t}) - \int_0^\tau \boldsymbol{W_{1hi}}(t) \boldsymbol{H_i}(\boldsymbol{t}) e^{\boldsymbol{x}_{2i}(\boldsymbol{t})'\boldsymbol{\beta_2} + \gamma_1(u_{0i} + u_{1i} * t)} dA_0(t) \right\} \right] \\
&= \sum_{i=1}^n \int_0^\tau E_{\boldsymbol{W_1}|Y}[\boldsymbol{W_{1hi}}(\boldsymbol{t})] d\boldsymbol{M_i}(\boldsymbol{t})
\end{aligned}$$

where

$$\boldsymbol{M_i}(\boldsymbol{t}) = \boldsymbol{N_i}(\boldsymbol{t}) - \boldsymbol{\Lambda_i}(\boldsymbol{t}) = \boldsymbol{N_i}(\boldsymbol{t}) - \int_0^t \boldsymbol{H_i}(\boldsymbol{u}) e^{\boldsymbol{x}_{2i}(\boldsymbol{u})'\boldsymbol{\beta_2} + \gamma_1(u_{0i} + u_{1i} * t)} dA_0(u)$$

is the usual counting process martingale for the i^{th} subject and the expectation is with respect to the conditional distribution of the random effects given the corresponding longitudinal measurements. Now we consider $(u_{0i} + u_{1i} * t)$ to be predictable and the variance of U to be estimated by

$$V_1 = \sum_{i=1}^n \int_0^\tau E_{\boldsymbol{W_1}|Y}[\boldsymbol{W_{1hi}}(\boldsymbol{t})]^2 d\boldsymbol{\Lambda_i}(\boldsymbol{t}) \tag{5.9}$$

Under mild conditions with independency between individuals, the martingale central limit theorem follows asymptotically with the implication that $U(\tau)/\sqrt{V_1(\tau)} \sim N(0,1)$ under H_0 as $n \to \infty$(Andersen et al., 1993).

5.4 Simulation studies

We carry out the simulation studies in three different categories.

5.4.1 Simulation study I - investigation of the power of the separate effect score test

The first simulation study investigates the power of the proposed score test for the separate association of the random effects. We generated data by using `simjoint()` function for Model B and we generated data by using our extended data generation function for Models C and D as explained in section 4.4. The key difference is `sepassoc` value in the function should be "TRUE" this time, which allows different values of γ for each random effect in the model . We investigated simulations in three different scenarios as we performed previously.

Each model and each scenario includes 100 simulations with different number of association parameters for $n = 500$ subjects. The association parameter, γ is chosen to be between 0 and 0.5 in increments of 0.03 (16 different values of γ) for Models B and C (as they both have two values of γ). As Model D has 4 random effects we have to choose 4 different values of γ to be able to achieve this simulation in a reasonable time frame. Therefore the association parameter is chosen to be $(0, 0.17, 0.33, 0.50)$ and we performed this simulation for all combinations of the supplied vectors. We performed this simulation $16^2 = 256$ different association pairs for Models B and C and $4^4 = 256$ different association combinations for Model D. Combined with the 100 simulations this gives rise to a large computational time.

These simulations were further complicated because of the bootstrap variance calculation. We again performed all these simulations in this chapter by remotely connecting to the Cauchy server. While the simulations of Models B and C took around 3.5 days for each scenario, the simulation of Model D took 26 days. Models B and C are run through all three scenarios; however, Model D is run only for Scenario III due to computational burden.

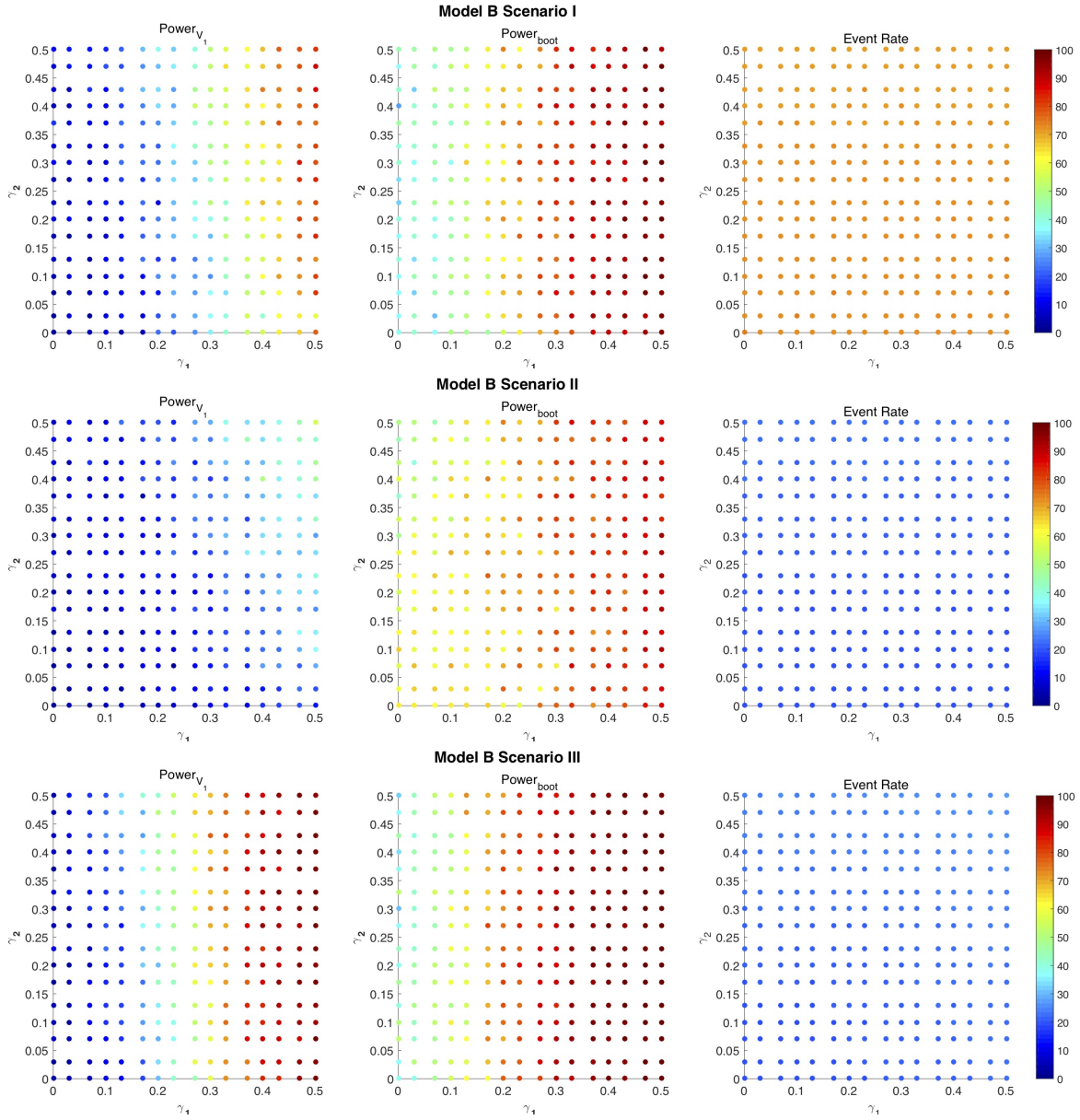

Figure 5.1: Power based on theoretical variance (left panel), power based on bootstrap variance (mid panel) and event rate (right panel) vs two different association parameters for Model B. The upper plots represent Scenario I, the middle plots represent Scenario II, the lower plots represent Scenario III.

Figure 5.1 indicates the plots of event rates and the powers of the score test based on the two variance estimators: martingale and bootstrap variance estimators for all three scenarios of Model B. Table 5.1 gives results regarding the powers and event rates from a representative selection of association pairs. We assume that the power is symmetric around 0 from the empirical results in the previous chapter. Therefore, we run the simulations for only positive values of γ pairs to alleviate the computational cost. Even though we have tested different association parameters for each random effect, the score test is still based on the seperate analysis of longitudinal and survival outcomes, analysed with `lme()` function in `nlme` package (Pinheiro et al., 2017), and `coxph()` function in `survival` package (Therneau, 2015), respectively. The results can be summarised as follows:

- As the association parameter for time-dependent random effect (γ_2) is constant and association parameter for time-independent random effect increases (γ_1) , power of the test increases quickly. Contrary to this, as the association of time-independent random effect (γ_1) is constant and the association of the time-dependent random effect increases (γ_2), the change of power is quite slow. It may be due to the fact that the test does not catch the association of the time-dependent random effects as easy as the one time-independent random effects. Another reason might be that some individuals might drop out very early and some might remain in the study for a long time, and have lots of measurements. The balance of the data is another issue that needs to be scrutinized here.

- For all scenarios, the power of the test based on bootstrap variance estimator is always higher than the power based on Martingale variance estimator. The reason might be for this is that the bootstrap variances are smaller compared to the theoretical variance; this may be an underestimate due to the small variances of random effects parameters.

- We experienced the same kind of pattern with the graphs in Chapter 4, Figure 4.2. The most powerful scenario is Scenario III, the least powerful

	$Power_{V_1}$	$Power_{boot}$	Event Rate(%)	γ_1	γ_2
Scenario I	4	42	73.330	0.030	0
	7	40	73.152	0	0.130
	79	96	72.686	0.500	0.230
	76	92	71.640	0.500	0.370
	79	99	70.854	0.470	0.500
Scenario II	5	63	19.696	0.030	0
	3	61	19.696	0	0.130
	40	91	20.298	0.500	0.230
	33	82	21.618	0.500	0.370
	45	88	22.836	0.470	0.500
Scenario III	6	39	19.586	0.030	0
	5	40	19.824	0	0.130
	97	100	23.800	0.500	0.230
	98	100	25.396	0.500	0.370
	97	100	26.408	0.470	0.500

Table 5.1: Indicative powers of the score test (in percent) for Model B with a representative small sample of the simulation results for each scenario. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

scenario is Scenario II.

- Event rates are pretty much constant as the magnitude of the association parameters, meaning powers are broadly comparable increase.
- Overall, the plots in Figure 5.1 demonstrate that the power increases as one of the γ_is, $i = 1, 2$ moves away from 0, as expected. Due to the differing nature of intercepts and slopes (the latter being stochastic in nature) results cannot be directly compared across scenarios.

Figure 5.2 displays the event rates and the powers of the score test with two variance estimators for Model C. Table 5.2 includes a small sample of the results given in Figure 5.2. For Model C, there seems to be not much difference between the performances of the test with the two variance estimators for all scenarios. In addition to this, the test seems to have less power compared to Model B up until $\gamma_i = 0.5$, $i = 1, 2$ for Scenarios I and II. Nevertheless, the power of the test looks like what we expect to see for Scenario III from our pre-

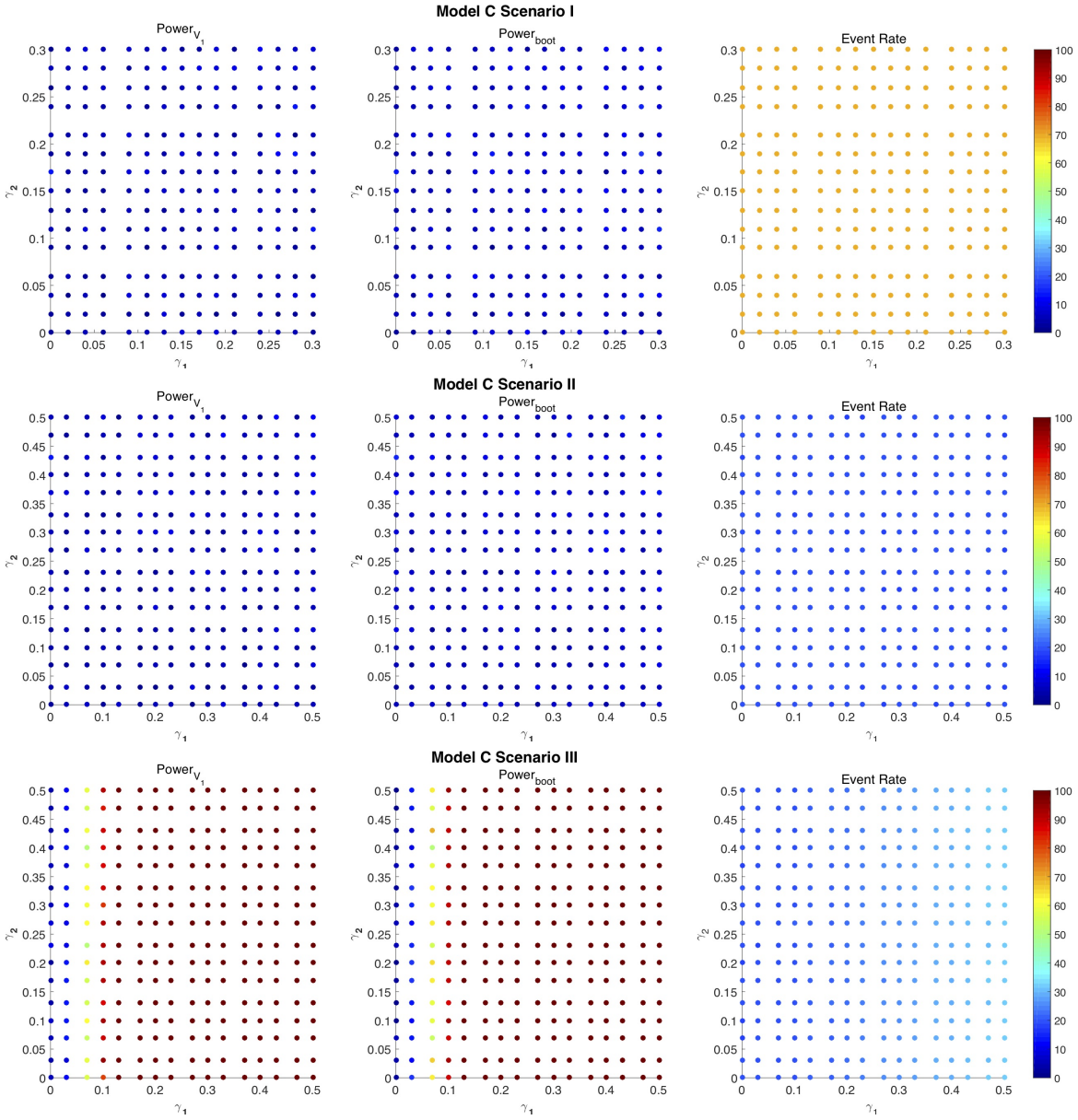

Figure 5.2: Power based on theoretical variance (left panel), power based on bootstrap variance (mid panel) and event rate (right panel) vs two different association parameters for Model C. The upper plots represent Scenario I, the middle plots represent Scenario II, the lower plots represent Scenario III.

	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)	γ_1	γ_2
Scenario I	2	3	69.546	0.020	0
	6	5	70.072	0.090	0.090
	6	10	69.750	0.150	0.170
	8	11	69.604	0.240	0.260
	7	8	70.032	0.280	0.300
Scenario II	5	7	18.792	0.030	0
	3	10	18.840	0	0.130
	6	7	19.030	0.500	0.230
	12	14	18.840	0.500	0.370
	3	3	18.726	0.470	0.500
Scenario III	12	16	19.302	0.030	0
	4	1	19.076	0	0.130
	100	100	32.224	0.500	0.230
	100	100	32.192	0.500	0.370
	100	100	31.274	0.470	0.500

Table 5.2: Indication of power of the score test in percentage for Model C with a representative small sample of the simulation results for each scenario. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

vious simulation studies in Chapter 4. In a narrower range, the test becomes quite powerful, due to the chosen parameters. The chosen parameters for each scenario are given in the simulation section, Chapter 4. Event rates are again reasonably constant for Scenario I and II and increase as the values of γ_i increase for the last scenario.

Table 5.3 gives some results of the powers and event rate of Model D for Scenario III. The sample selected randomly and we put results when $\gamma = 0.330$ for each random to be able to make a comparison of the performances of the random effects regarding time included or not. We are not able to plot the power of the test for Model D due to the number of random effects. There are 4 random effects in Model D and each random effect has its own association parameter. We can say that as the association parameter of time-dependent random effect is bigger than 0 and the other three are 0, the test is less powerful compared to the one with time excluded random effects. The test gets powerful quite

	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)	γ_1	γ_2	γ_3	γ_4
Scenario III	3	2	21.130	0	0	0	0
	5	5	22.504	0	0.170	0	0
	100	100	25.294	0.500	0.170	0.170	0
	85	81	27.132	0	0.500	0.330	0
	96	97	28.034	0.170	0.500	0	0.170
	66	52	24.364	0	0.170	0.330	0.170
	100	100	30.748	0.500	0.500	0.330	0.330
	100	100	30.258	0.170	0.500	0.500	0.330
	79	83	30.896	0.170	0.330	0	0.500
	99	100	30.638	0.330	0.170	0.330	0.500
Scenario III	96	96	22.556	0.330	0	0	0
	17	22	24.240	0	0.330	0	0
	74	58	22.162	0	0	0.330	0
	36	15	26.010	0	0	0	0.330

Table 5.3: Indication of power of the score test in percentage for Model D with a representative small sample of the simulation results for Scenario III. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

quickly in this scenario. For the whole results, see Appendix A.3.

5.4.2 Simulation study II - investigation of the power of the multivariate score test

The background methodology of the data simulation, the analysis of the simulated data and the simulation results of the score test for the multivariate joint data are given and discussed in this section.

In this simulation study, the joint data containing three continuous normally distributed longitudinal outcomes and a single survival outcome are simulated using `simData()` function in `joineRML` in `R` (Hickey et al., 2017). Each simulated dataset contains 250 subjects. A maximum of 8 longitudinal measurements at times 0,1,2,3,4,5,6,7 are permitted. Measurements are taken up to the event time for each subject.

The longitudinal outcomes are simulated under a model including a combination of fixed and subject-specific random effects with an independent and

identically distributed error term. The fixed effects are time, a continuous variable, and a binary indicator randomised equally to two arms, along with an intercept term. Random effects are subject level only, consisting of either random intercept only or random intercept and time (slope) terms. The individual level random effects follow a multivariate normal distribution and they are independent of error terms. As always in this work, the random effects link the longitudinal and survival submodel with association parameters.

Event times are simulated in the same way as explained earlier in Section 4.4. The key difference here is the latent association of time-to-event submodel is a linear combination of latent associations in the longitudinal submodels. The event rate is tuned through the baseline hazard parameters θ_0 and θ_1, for fixed values of the other parameters.

In the longitudinal submodels the fixed effects regression coefficients are set to

$$\boldsymbol{\beta} = \begin{bmatrix} 0.500 & 2 & 1 & 1 \\ 2 & 2 & -0.500 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Each row of the coefficients corresponds to the related longitudinal biomarker in the order of an intercept, a time, a continuous and a binary covariates, $\beta_2 = (1, 1)$ specifyies the coefficients for the time-to-event baseline covariates, in the order of a continuous and a binary covariates. The subject level random effects are generated under a standard normal distribution as default. The measurement errors are generated under a normal distribution with mean 0 and standard deviation 0.25 for each biomarker, fixed so as not to add another layer of uncertainty.

In order to investigate the power of the test with the magnitude of the association parameter, we chose 25 association levels from 0 to 0.5 in increments of 0.02 for the first biomarker, while the second and third ones are fixed as $\gamma_2 = 0.9$ and $\gamma_3 = 0.5$, respectively. Only positive associations are examined, as the behaviour of the power of the test for negative association is expected to be

similar from our empirical results in the previous chapter.

The simulations are repeated 100 times for all 25 γ_1 parameter without bootstrapping. Despite this, the convergence was quite slow. It took 4.5 days to complete the whole simulation.

Following the method introduced in Section 5.3.2, the test is based on the bivariate model fitting with `mjoint()` function to each dataset excluding the biomarker that we hypothesise not to affect survival (Hickey et al., 2017). The model is set to either random intercept or random intercept and slope model at the subject level. The association structure linking the submodels is a proportional random effects structure. The random effects and the coefficients of the submodels regarding the two longitudinal biomarkers and time-to-event are extracted from the bivariate model fit. We also fit a multivariate linear mixed effect model for all the biomarkers accounting for the correlations amongst them with a function called `mvlme()`, which is an internal function in `mjoint()` function. The details of this function can be found at `https://github.com/graemeleehickey/joineRML/blob/master/R/mvlme.R`. However, we modified this function to be able to extract the individual random effects for each biomarker. Then the code works on the baseline hazard function at each time point for each subject for counting process. Lastly, we extract the random effects of the longitudinal biomarker that we tested from the multivariate linear mixed effect model fit to calculate the score statistics. The modified function can then be found at `https://github.com/goncabuyrukoglu`.

Results from the simulation study for the power of the score test of multivariate joint data are presented in Figure 5.3 for the random intercept model and random intercept and slope model, respectively. We will refer to the multivariate joint model with random intercept as Model E, and that with random intercept and slope as Model F. Table 5.4 represents a small sample of result of the simulation study for each model. As expected, the power of the test increases as the magnitude of the corresponding association parameter gets bigger for random

	γ	$Power_{V_1}$	Event Rate(Percentage)
Model E	0.080	15	63.144
	0.190	39	63.312
	0.290	82	63.056
	0.400	99	62.936
	0.500	100	63.128
Model F	0	6	64.744
	0.110	7	63.896
	0.220	16	63.728
	0.330	31	63.912
	0.440	38	63.768

Table 5.4: Indication of power of the score test in percentage for Models E and F with a representative small sample of the simulation results. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

intercept model. The behaviour of the power of the test for Model F is similar but with a slower rate. To be able to get 100% power of the test we need higher magnitude of association parameter with the same scenario. It may be due to the fact that the latter model includes time, and time might not be well-balanced for every individual in the simulation. Another reason might be the increased dimensionality of the random effects components making the model struggle to catch the expected results easily. The event rates for both models are pretty much constant. The mean of event rate for Model E is 63.10% and for Model F is 63.87%, the simulation time for Model E is 4.5 days for 25 value of γ, for Model F is 3 days for 10 values of γ. Due to the computational burden of the calculation for Model D, the simulation investigates fewer γ values than Model C. We expect the results to behave similarly for the other choice of γ.

5.4.3 Simulation study III - investigation of the power of the multilevel score test

The data simulation for multilevel joint data both types of model at both levels are explained in detail in Section 3.5.1. We generate data with Models C and D for this section, which are explained earlier. In each setting, we generate 25

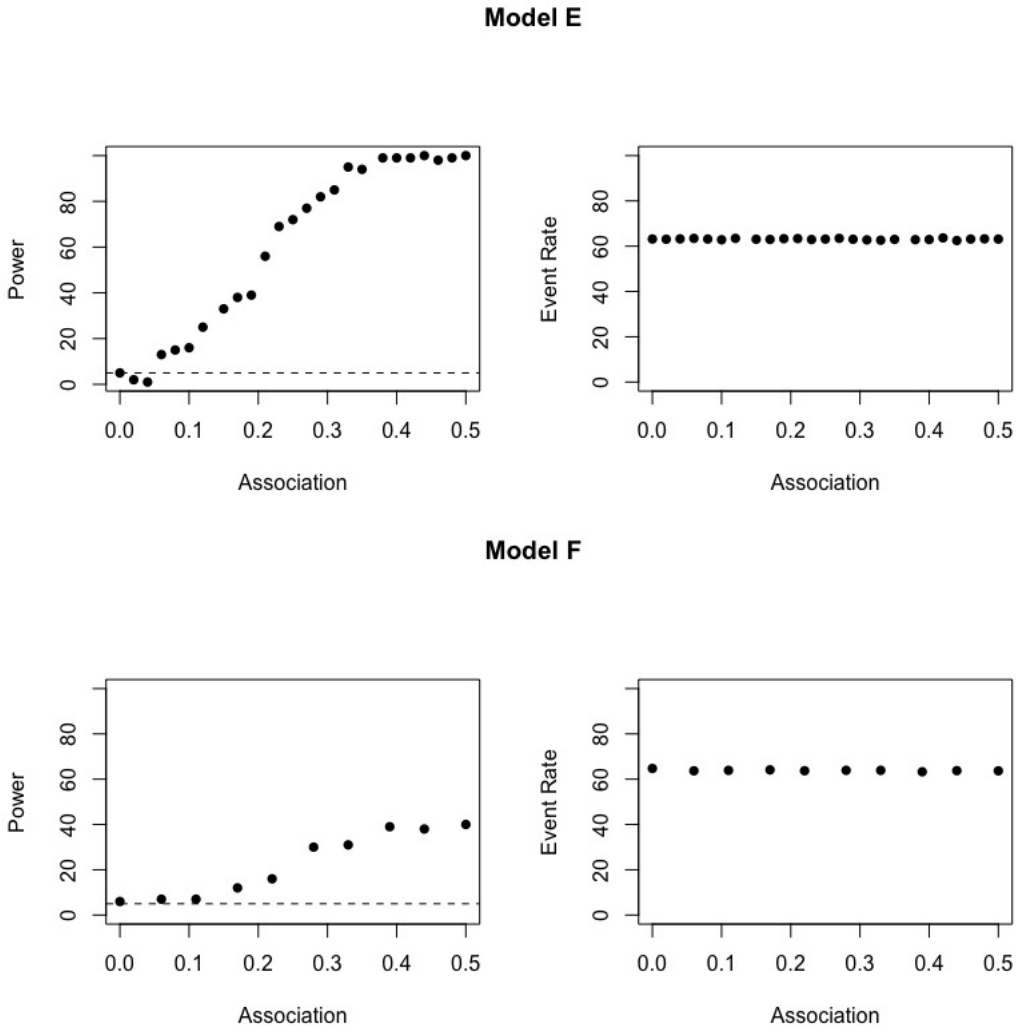

Figure 5.3: Powers of the multivariate score test and event rates of the simulated dataset for Models E and F.

subjects ($m = 25$) in each of 20 centres ($noh = 20$), totally $n = 500$ subjects. Each simulation is repeated 100 times. The means event rate of Models C and D are 48.20% and 48.26%, respectively. The simulations took 17 hours and 3.5 days for 50 gamma value for Models C and D, respectively.

The simulation results in terms of the power of the score test for multilevel joint data for both model and the event rates of the simulated datasets are presented in Figure 5.4 and a small representative sample is shown in Table 5.5. The γ_1 value for both model is chosen as 0.4. The shape of the power of the test for Model D is noisy but it is what we would expect to see. Nonetheless, we come across an unexpected pattern of the shape of the power of the test for Model C. We would expect to an increasing pattern for positive γ values because it is symmetric around 0 in the previous scenarios. However, in this graph, it is symmetric around 0.4, which is chosen value of γ_1 for subject level in the joint model. This requires further investigation since the nominal power under the null should be 5%. Using a larger sample and letting number of center to be 1 should allow for a fuller understanding but is the beyond the scope of this thesis.

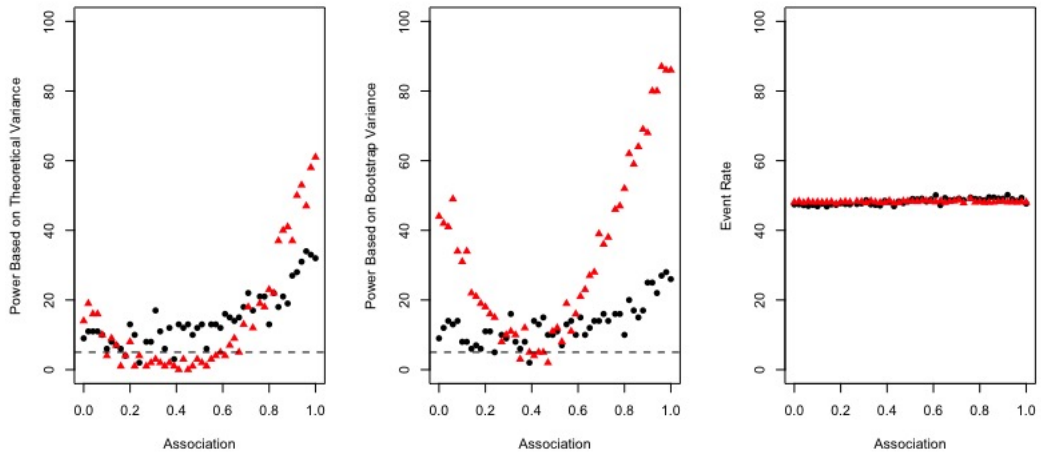

Figure 5.4: Power of multilevel score test and event rate:● (black) for random intercept and slope model at both level; ▲ (red) for random intercept model at both level. Theoretical variance (left panel), bootstrap variance (mid panel), event rate (right panel).

	γ_1	γ_2	$Power_{V_1}$	$Power_{boot}$	Event Rate(Percentage)
Model C	0.400	0	14	44	48.084
	0.400	0.180	4	19	47.638
	0.400	0.390	1	5	47.922
	0.400	0.590	5	16	48.322
	0.400	0.800	23	52	48.222
	0.400	1	61	86	48.064
Model D	0.400	0	9	9	47.432
	0.400	0.180	4	6	47.328
	0.400	0.390	3	2	48.486
	0.400	0.590	12	10	48.900
	0.400	0.800	13	10	48.914
	0.400	1	32	26	47.670

Table 5.5: Indication of power of the score test in percentage for Models C and D with a representative small sample of the simulation results. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

5.5 Application: the scleroderma lung study

We now return to the SLS data to apply the proposed method in the first category, which was also used to illustrate score test in the previous chapter. This study was designed to evaluate the effects of oral cyclophosphamide (CYC) versus placebo treatment in patients with evidence of active alveolitis and scleroderma-related interstitial lung disease. The repeated measurement FVC% was taken over a long period of time as the primary outcome. The measurements were taken every 3 months during 2 years. Missingness in FVC% can be nonignorable as they may be associated with the failures, which were adverse events and serious adverse events, deaths, dropout, and treatment failure, pooled according to the criteria explained in the first chapter if the events had correlation between unobserved FVC% measurements. In this section, we analyse the CYC treatment effect on FVC% by using a joint model consisting of the submodels (4.1) and (4.2).

We first apply this dataset to the proposed model with random intercept and

slope at subject level, Model B, Equation (4.4) , random intercept at both subject and centre level, Model C, Equation (3.3) and random intercept and slope at both subject and centre level, Model D, Equation (3.4) which are explained in Section 4.2 in conjuction with separate association. Now we consider the association parameter $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)$ where q is the number of random effects. That is, each γ_i is different from each other $(i = 1, \ldots, q)$. We would like to test if we need the association between two components of the joint model while each γ_i is different. Table 5.6 represents the results regarding the statistics to test the null hypothesis with martingale variance estimator and bootstrap variance estimators and corresponding p-values for Models B, C and D. As can be seen from this table, the only significant test statistic is the first one using the martingale variance estimator of Model B. The rest are nonsignificant. However, our simulation study shows the power of the test when this dataset is mimiced for these models.

Table 5.7 presents parameter estimates, their standard errors obtained bootstrap variance and log-likelihoods for random intercept and slope model, Model B. The association parameter estimates are $\gamma_B = \{-0.455(1.679), -0.065(0.129)\}$. We would like to find out how reliable the test results are from the Simulation Study I. We chose the values for γ as $\{0.47, 0.07\}$ with the event rate 22.28 based on our experience from the simulation studies in Chapter 4 indicating that the power of the test is symmetric around zero. The power of the test based on theoretical variance is 88%, the power of the test based on bootstrap variance is 97%. However, overall, the bootstrap variance seems underestimated. Therefore we only consider power based on martingale variance estimator. So we can say that for this scenario the test is quite powerful and we can rely on these results and say that there is need to be fitted joint model with random intercept and slope at subject level in conjunction with separate association. The separate association fitting for Models C and D have not been implemented yet. So, we can leave them as future work for now.

	χ^2-value (Martingale Variance)	p_{V1}-value	χ^2-value (Bootstrap Variance)	p_{boot}-value
Model B (DF)	14.626 (2)*	0.001	3.718 (2)	0.156
Model C (DF)	1.551(2)	0.460	2.253(2)	0.324
Model D(DF)	3.028(4)	0.553	2.715(4)	0.606

Table 5.6: Score test statistics and p-values of the sls data under the null hypothesis using martingale and bootstrap variance having separate association between the two components.

	Model B	
	$W_1(t) = U_1 + U_2 t$, $W_2(t) = \gamma_1 U_1 + \gamma_2 U_2 t$	
	Est	SE
Longitudinal		
Constant	0.799	2.796
MAXFIB	-0.150	0.473
CYC	6.017	3.897
Months	-0.027	0.068
FVC0	0.976	0.037
MAXFIB:CYC	-1.018	0.555
CYC:Months	-0.112	0.088
CYC:FVC0	-0.051	0.055
Survival		
MAXFIB	0.322	0.215
CYC	-0.065	0.431
Association		
γ_1	**-0.455**	1.679
γ_2	**-0.065**	0.129
Random effects		
σ^2_{u1}	3.176	1.642
σ^2_{u2}	0.193	0.044
Noise		
σ^2_{ε}	23.870	4.295
Log likelihoods		
Longitudinal	-3366.958	
Survival	-115.1238	
Combined	-3482.082	

Table 5.7: SLS data results for random-intercept-and-slope with separate association (Model B)

5.6 Discussion

In this chapter, we deal with a longitudinal biomarker, multiple-correlated longitudinal biomarkers and a multilevel (hierarchical) longitudinal biomarker as growth curves and extend the method proposed by Henderson et al. (2002) and examined in Chapter 4 to determine if the corresponding longitudinal biomarker or centre level effect in the longitudinal biomarker is significantly associated with event times data. To do this analysis, our assumption is that random effects in the growth curve ($\boldsymbol{W_{1i}(t)}$) are predictable. The derivation of a score test is necessary to help to decide whether the two components are associated and whether a joint model or more complex joint model is fitted. While the test is based on the separate analysis of the two components of the model and employs a linear mixed effect model and a Cox PH model in the first category of the extensions of the score test in this chapter, the second and the third extensions still remain to a multivariate joint model and a univariate joint model, respectively, as a surrogate for a time-to-event outcome, and a Cox PH model.

We conduct several simulation studies to examine the power of the score test for these three models. The simulation of the first models are carried out for three scenarios as also examined in the previous chapter. The multivariate score test and the multilevel score test are examined for one scenario due to their intensive model fit. The results of the score test for separate effect association and multivariate score test indicate that the test is more powerful under time-independent random effects model structures, i.e., random intercept model. When the model becomes time-dependent, the test struggles to identify the association between the two components. The parameters that are chosen for the generation of the data have also effects on the power of the test. For example, the test becomes powerful quickly compared to Scenario I and II in the examination of score test for separate effect association for Scenario III, as it is also seen in our empirical results in the previous chapter.

We found an unexpected result from Simulation Study III that the test depends heavily on the value of γ for subject level random effects for random intercept model. We still have the right shape as we expect to see but it is a location shift. The shift is correspond to the size of γ for the subject level. The power of the test for random intercept and slope model is quite noisy; however, we can see an increasing pattern as γ values move away from 0 as expected.

We have an application of the SLS data to our first category of the extensions of the score test. The results demonstrate that only the association for individual level random intercept and slope model is significant based on Martingale variance estimator. This results contradicts based on the bootstrap variance estimator. Having consideration of our empirical results in Section 5.4.1, the bootstrap variances might be underestimated and this makes the test so powerful at this scenario. The rest of the models do not find that the association is statistically significant. This is most likely to be explained by the fact that the dataset has fewer number of the centres than our simulation scenarios and the number of individuals for each centre is not equal. The dataset is not well-balanced in terms of spread of the sample size to the centres. While some centres have less than 3 subjects, some centres have more than 15 subjects. However, in our simulation study, we choose the equal number of individual for each centre.

The investigation of the power of the test for all three categories with bigger or smaller sample size with unequal sample size in each centre might be a possible future work of this study. Having more than three correlated longitudinal biomarkers with different scenarios and event rates can be another further work for this chapter.

Chapter 6

Joint Modelling of Time-To-Event and Multivariate Longitudinal Data: An Application to the ADNI Dataset

6.1 Introduction

In Chapter 5, we derived a suite of score tests to identify the relationship between univariate longitudinal and survival data in conjunction with separate association of random effects, a score test to determine if higher levels of random effects have an association in the presence of hierarchical or clustered data between a longitudinal biomarker and time-to-event data, and a score test to detect the association between multiple longitudinal biomarkers and survival times.

In this chapter, a joint model with multiple longitudinal and a single survival outcomes is defined. This joint modelling structure is an extension of the ubiquitous joint model of single longitudinal and single time-to-event data to the multivariate case of multiple longitudinal outcomes. The model provides an effective tool that fulfils several of the desired features of the needed new risk prediction models (Hickey et al., 2016), and is extended by Lin et al. (2002).

Despite the fact that univariate joint modelling has seen an increase in clinical

application, there are only a handful of statistical software packages available to analyse data, and the models are more complex than those used for a single longitudinal outcome and a single event time. For example, Rizopoulos and Ghosh (2011), motivated by a real dataset on renal graft failure, proposed a new semiparametric joint model of multivariate longitudinal data and a single time-to-event. The dataset consisted of 407 patients with chronic kidney disease following a renal transplantation. They concentrated on three longitudinal biomarkers known to be related to graft functioning and graft survival, which are the glomerular filtration rate (GFR, continuous), the proteinuria (binary), and the blood haematocrit level (continuous). Putting all of the available information in one model should make the prediction more efficient, more powerful, and use the data optimally. This highlights the need for joint models with multivariate longitudinal outcomes and a time-to-event outcome to optimally capture the information for the underlying disease outcome, and ultimately offer the most optimal treatment for the patient.

Simultaneous modelling of multiple longitudinal biomarkers and event time data in joint models offers a number of advantages over separate joint modelling of each biomarker (Rizopoulos and Ghosh, 2011; Brown et al., 2005; Chi and Ibrahim, 2006; Xu and Zeger, 2001; Song et al., 2002a). The first advantage, as stated by Rizopoulos and Ghosh (2011), is that it is more accurate to estimate the association after having adjustment for the effects of others for correlated longitudinal biomarkers. Secondly, Fieuws et al. (2007) indicated that taking into account the correlation between longitudinal biomarkers may substantially improve the predictive ability of a joint model. Furthermore, accommodation of association between longitudinal biomarkers has been shown to be more efficient when compared with modelling each of them separately in some settings (McCulloch, 2008; Gueorguieva and Sanacora, 2006). Hickey et al. (2018) showed that fitting a univariate model inflates the association between biomarker and the event, akin to omitting a covariate in a regression model.

Individualized predictions for the survival and longitudinal outcomes have dynamic nature and they can be obtained from fitted joint models. As time progresses, more information is recorded for the individual, and thus the predictions can be updated based on the new information. Rizopoulos (2011) presented dynamic survival probabilities of a individual based on the available longitudinal history data, derived accuracy measures under a joint modelling framework, and assessed the discrimination capability of a marker between subjects who had the event within a short time from subjects who did not. In this chapter, the dynamic survival probability and the conditional expected value of a longitudinal biomarker for a new subject are investigated from the last observation time, given the longitudinal trajectories of the subject. The aim is also to gauge if predictive ability is improved by fitting a multivariate joint model over fitting univariate joint models separately.

The remainder of this chapter is organized as follows: patient population and the study's design are explained briefly in Section 6.2. The subsequent section explains the measures that are considered in the model. In Section 6.4, the methodology is applied to the ADNI dataset to assess the association between various multiple longitudinal biomarkers and time-to-dementia. Section 6.5 illustrates the estimation of survival probabilities in joint modelling framework, and presents predictive accuracy for repeated measurements, whereas Section 6.6 provides the results of the multivariate joint models as well as the univariate joint models, along with the score test calculations, and finds out the discrimination capability of the markers. The final section reflects on the findings of the chapter.

6.2 Patient population and study design

The ADNI dataset, introduced in Section 1.5.3, is the focus of the joint models with a solitary time-to-event and multivariate longitudinal data. It investigates the progression of Alzheimer Disease (AD) using serial magnetic resonance imaging (MRI), positron emission tomography (PET), some serum biomarkers,

and various clinical and neurocognitive measures. For convenience, each of these longitudinal measures are referred to as "biomarkers". A total of 388 patients with amnestic mild cognitive impairment (MCI) at baseline with at least one follow-up visit were included in the analysis.

6.3 Measures

This study collected data from a wide range of clinical and biological records of patients. The dataset included 33 different biomarkers which are considered to be candidates to help detect the conversion of MCI to AD. The measurements were taken from the patients in the ADNI-I study at baseline, 6, 12, 18, 24, and 36 months, and annually thereafter. There are three domains considered for this study: neuropsychological assessment (8), functional and behavioural assessment (15), and neuroimaging (10). Each of these domains is explained below.

6.3.1 Neuropsychological assessment

The neuropsychological assessment domain includes measurements of Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog), which assesses patients' written and verbal responses that are related to essential cognitive function, with scores from 11 items, or, in the expanded version, 13 items, and a score range between 0 and 70 or 0 and 85 (expanded). The higher the score, the poorer the cognitive function. The rest of the measurements in this domain assessed the verbal memory (the Rey Auditory Verbal Test: RAVLT immediate, RAVLT learning, RAVLT forgetting), as well as conducted assessments such as the Mini-Mental State Examination (MMSE, a total of 11 questions and a score range between 0 and 30, where the lower the score, the more severe the cognitive impairment), the Montreal Cognitive Assessment (MOCA), a 30 point test that assesses various cognitive domains, and the Clinical Dementia Rating Sum of Boxes (CDR-SB, assessing the stage of severity of dementia with a score range from 0 to 18). These measurements are summed scores, and as such, they are considered to be continuous.

6.3.2 Functional and behavioural assessment

Measurements in the functional and behavioural assessment domain involved the Functional Assessment Questionnaire (FAQ; a total of 10 items with scores ranging from 0 to 30 - where higher scores reflect greater functional dependence), Everyday Cognition by the Patient (ECogPt), and Everyday Cognition by the Patient's Study Partner (ECogSP), where the latter two assess the patient's ability to carry out normal everyday tasks in multiple domains (memory, language, visuospatial abilities, planning, organization, and divided attention).

6.3.3 Neuroimaging

The data in the neuroimaging domain was used in several previous studies (Landau et al., 2011; Jagust et al., 2010; Landau et al., 2012; Fennema-Notestine et al., 2009). Measurements from PET imaging are a sum of mean glucose metabolism uptake in regions of angular (right and left), temporal (right and left), and posterior cingulate (FDG-PET) (Landau et al., 2011), the average of standardized uptake value (SUVR) of frontal cortex, anterior cingulate, precuneus cortex, and parietal cortex (PIB) (Jagust et al., 2010), and the average of florbetapir SUVR of frontal, anterior and posterior cingulate, lateral parietal, and lateral temporal cortex (AV45) (Landau et al., 2012). On the other hand, measurements from MRI imaging are volumetric data of ventricles, hippocampus, whole brain, entorhinal, fusiform gyrus, middle temporal gyrus and intracerebral volume. These data are collected with 1.5T or 3T MRI scanners, and volumes of the region of interests (ROIs) were reconstructed with the *Freesurfer* software (Fennema-Notestine et al., 2009).

6.4 Statistical models

Data from the multiple biomarkers were collected from the patients at different time points during the follow-up period. The hypothesis tested here is that the corresponding biomarkers are associated with the time-to-dementia - the event of interest. First, the time-to-dementia was modelled together with multiple

longitudinal trajectories for the aforementioned variables from three domains using univariate and multivariate joint models. The joint model, regardless of whether univariate or multivariate in nature, consists of two submodels: the longitudinal and the survival submodels. These submodels are explained in Section 2.4 for the univariate case, and Section 5.2 for the multivariate case. The longitudinal submodel allows one to characterize the change in trajectories over time with the adjustment of age at baseline and the presence of apolipoprotein E (APOE) $\varepsilon 4$ allele. The random effects model included in the submodels is a random intercept and slope model (which seems reasonable based on the trajectories of the biomarkers from Figures 6.1 and 6.2). The event time was defined from the baseline until the event happens (conversion from MCI to AD or censoring). The measurements continued to be taken after conversion to AD for some participants. However, such measurements were excluded from the analysis. The baseline covariates in the survival submodel are gender, age at baseline, presence of APOE $\varepsilon 4$ allele, and years of education. An association parameter linked these submodels, assuming that the hazard is affected by the longitudinal measurements via the current value.

Suppose $Y_{ik}(t)$ is an $(n_{ik} \times 1)$-vector of observed longitudinal measurements for the k^{th} biomarker: $Y_{ik} = (Y_{i1k}, \ldots, Y_{in_{ik}k})'$ at time t for the subject i, $i = 1, \ldots, n$. Subject i has n_{ik} number of measurements for the k^{th} biomarker. The longitudinal submodel for multivariate biomarkers is represented as:

$$Y_{ik}(t) = \boldsymbol{x}'_{1ik}\boldsymbol{\beta_{1k}} + \boldsymbol{W}_{1i}^{(k)}(t) + \varepsilon_{ik} \tag{6.1}$$

where $\boldsymbol{x}'_{1ik} = (1_{ik}, Age_{ik}, I(APOE\varepsilon 4_{ik} > 0))$, is an $(n_{ik} \times 3)$-matrix of observed longitudinal measurements of the i^{th} subject for the k^{th} biomarker, $\boldsymbol{\beta}_{1ik}$ is the corresponding fixed effect terms, $\boldsymbol{W}_{1i}^{(k)}(t) = \boldsymbol{D}'_{ik}(t)\boldsymbol{U}_{ik} = U_{0ik} + U_{1ik} * t$ is the random intercept and slope model which indicates subject-specific change rate of the measurement, $\boldsymbol{D}_{ik}(t)$ is an r_k-vector of potentially time varying covariates, $\boldsymbol{U_{ik}}$ are random effect terms following a zero-mean multivariate normal distribution with (2×2)-variance-covariance matrix $\boldsymbol{\Sigma_{kk}}$, and, finally, $\varepsilon_{ik} \overset{i.i.d.}{\sim} N(0, \sigma_k^2)$ is the error term, all for $k = 1, 2, \cdots, K$.

The survival submodel is represented as follows:

$$\lambda_i(t) = \lambda_0(t)\exp\{\boldsymbol{x}'_{2i}\boldsymbol{\beta_2} + \boldsymbol{W_{2i}(t)}\}, \tag{6.2}$$

where $\lambda(\cdot)$ is an unspecified baseline hazard, $\boldsymbol{x}'_{2i} = (Gender_i, Age_i, Education_i, I(APOE\varepsilon 4_i > 0))$ is a $(n \times 4)$-matrix of covariates with corresponding vector of regression coefficients $\boldsymbol{\beta_2}$. Random effects in the hazard function in Equation (6.2) are specified as $\boldsymbol{W_{2i}(t)}$, which is a linear combination of $\{\boldsymbol{W_{1i}^{(1)}(t)}, \cdots, \boldsymbol{W_{1i}^{(K)}(t)}\}$:

$$\boldsymbol{W_{2i}(t)} = \sum_{k=1}^{K} \gamma_{yk} \boldsymbol{W_{1i}^{(k)}(t)},$$

where $\boldsymbol{\gamma_y} = (\gamma_{y1}, \cdots, \gamma_{yK})$ are the vector of joint model association's parameters. A significant γ_{yk} indicates that there is a strong association between longitudinal biomarkers and time-to-dementia. Each biomarker is scaled to zero-mean and unit variance using the mean and standard deviation amongst all the individuals and the measurement times. This is done in order to be able to make a proper comparison of the γ_{yk} parameter values among biomarkers.

Dropping subscript k, the multivariate model reduces to a univariate joint model. The univariate joint model of 33 biomarkers were fitted seperately in Li et al. (2017) using a Weibull parametric baseline hazard. However, our univariate joint models are slightly different than those fitted in that paper owing to our choice of hazard. We associated longitudinal and survival submodels through the random effects only, while they are associated with these submodels through random effects and fixed effects, which is commonly known as the "current value" approach.

The strongest seven predictors, as specified by Li et al. (2017), were chosen to fit multivariate joint models. They are ADAS-Cog 13, ADAS-Cog 11, RAVLT Immediate and RAVLT Learning in the cognitive domain, FAQ in the functional domain, middle temporal gyrus (MidTemp), and hippocampal volume in the neuroimaging domain. However, the correlation between ADAS-Cog 13 and ADAS-Cog 11 is found to be 0.96, indicating that they are highly correlated as one might expect. As such, excluding ADAS-Cog 11, the models were fitted

with the rest of the six biomarkers. The correlation matrix of the rest of the biomarkers is given in Table 6.1. The table shows that the correlations between biomarkers do not seem high, hence, the models including a combination of these biomarkers may be expected to be representative models.

	ADAS13	RAVLT.imm	FAQ	Mid	RAVLT.learn	Hip
ADAS13	1	-0.670	0.399	-0.355	-0.459	-0.447
RAVLT.imm	-0.670	1	-0.288	0.242	0.579	0.312
FAQ	0.399	-0.288	1	-0.197	-0.243	-0.246
Mid	-0.355	0.242	-0.197	1	0.142	0.559
RAVLT.learn	-0.459	0.579	-0.243	0.142	1	0.294
Hip	-0.447	0.312	-0.246	0.559	0.294	1

Table 6.1: The correlation matrix of the six biomarkers of the ADNI data.

If the biomarkers are very highly correlated, one may expect the model to struggle to identify all association parameters. Note that the univariate joint model does not have such an issue. However, as soon as we move towards models that have multiple longitudinal biomarkers, we must consider how related the biomarkers are to each other. Otherwise, a standard problem known as multicollinearity may arise. Multicollinearity is a common problem in statistics, where one variable in a multiple regression model can be linearly estimated from another with a substantial degree of accuracy. This may affect the coefficient estimates erratically in response to small changes in the model or the data. Furthermore, it may inflate the standard errors, and cause the coefficients to be driven towards being non-significant. All of these biomarkers measure cognitive decline, and as such, one may reasonably expect them to be related to Alzheimer's disease. Assessing the correlation amongst biomarkers and multicollinearity can be a preliminary analysis before fitting the multivariate joint models. If the models are independently fitted with each biomarker allowed to contribute, there is much to gain in using multivariate joint models.

With the trio combinations of the strongest six predictors (excluding ADAS-Cog 11), 20 multivariate joint models in total were fitted in Tables 6.4 and 6.5. These models were fitted using the `mjoint` function in `joineRML`, and γ, Z, and

p values are given for each biomarker in those tables.

The score test was also adapted to identify the association between the submodels for both univariate and multivariate joint models. The score statistics and the corresponding p values based on the theoretical variance and bootstrap variance estimators for univariate joint models are presented in Table 6.6. We aim to compare the score statistics based on two different variances as a prognostic tool by fitting the models with `joineR` and `joineRML` packages, along with a comparison of the Z value (calculated by the `joineRML` package) and the Z^* value (calculated by the `JM` package, taken from Li et al. (2017)) after fitting the models. The score test for the univariate joint model is explained in Section 4.3, while the score test for the multivariate joint model is explained in Section 5.3.2.

The score statistics and corresponding p values of each biomarker in each multivariate joint model are presented in Tables 6.4 and 6.5. The aim is to identify the need for the biomarker of interest given the presence of other biomarkers. As the model still remains a multivariate joint model (bivariate) while testing one of the biomarkers, the convergence of the models was quite slow. Therefore, no attempt to calculate bootstrap variance estimators was made here due to the computational burdensome.

6.5 Predictive survival probabilities and prospective accuracy for the multivariate joint models

In this section, the dynamic predictions of survival probabilities and the discrimination capability of a longitudinal marker are assessed.

6.5.1 Predictive survival probabilities

We also focus on dynamic predictions for the both submodels, based on an observed measurement history for the repeated outcomes of a subject, using

the Monte Carlo simulation method described by Rizopoulos (2011). Survival probabilities for a new subject i, who has survived up to time t, and has had repeated measurements $Y_{ik} = \{y_{ik}(t_{il}); 0 \leq t_{il} < t, l = 1, \ldots, n_i\}$ are derived. The survival probabilities at time u are:

$$\pi_i(u|t) = Pr\{T_i^* > u | T_i^* > t, Y_i(t), \omega_n\} \tag{6.3}$$

$$= \int \int \frac{S\{s|H_i(s, u_i), \theta\}}{S\{t|H_i(t, u_i), \theta\}} f\{u_i|T_i^* > t, \boldsymbol{Y_i}(t), \theta\} f(\theta|\omega_n) du_i d\theta \tag{6.4}$$

where $\omega_n = \{T_i, \delta_i, y_i; i = 1, \ldots n\}$ denotes a sample from the target population, $S(\cdot)$ is the survival function ($\boldsymbol{Y_i}(t) = \{Y_{i1}(t), \ldots, Y_{iK}(t)\}$), and $\boldsymbol{H_i}(t)$ denotes the history of the whole longitudinal process. The following simulation scheme can obtain a Monte Carlo estimate of $\pi_i(u|t)$:

Step 1 . Sample a value $\tilde{\theta}$ from the posterior of the parameters $[\theta|\omega_n]$.

Step 2 . Sample a value $\tilde{u_i}$ from the posterior of the random effects $[u_i|T_i^* > t, \boldsymbol{Y_i}(t), \theta]$.

Step 3 . Calculate the ratio of the survival probabilities $\frac{S\{s|H_i(s,\tilde{u_i}),\tilde{\theta}\}}{S\{t|H_i(t,\tilde{u_i}),\tilde{\theta}\}}$.

Step 4 . Repeat Steps 1-3 for each subject L times. The conditional survival probabilities can be estimated by:

$$\frac{1}{L} \sum_{l=1}^{L} \frac{S\{s|H_i(s, \tilde{u_i}^{(l)}), \tilde{\theta}^{(l)}\}}{S\{t|H_i(t, \tilde{u_i}^{(l)}), \tilde{\theta}^{(l)}\}}$$

and the standard error can be estimated by the standard deviation of the Monte Carlo samples.

6.5.2 Discrimination for multivariate longitudinal biomarkers

Standard comparison of likelihoods or information criterion (i.e. AIC) are not appropriate when comparing joint models with a differing number of biomarkers. The joint likelihood, as seen earlier in Section 2.6, can be split into two components: one from the longitudinal submodel, and the other from the survival submodel. The dimension of the longitudinal model changes when multiple biomarkers are considered, and the data themselves may change, since

the measurement schedule may vary between biomarkers. Furthermore, even in the case of a fixed schedule, not all biomarkers are necessarily measured at each time point, as is the case in the ADNI data. Looking at the significance, or otherwise, of association parameters in a multivariate joint model is informative to an extent, as it typically leads to attenuation towards the null for the association parameters from a univariate joint model. The issue of model performance is instead investigated by two main arms, where the first one involves calibration measures and the assessment of performance of the model prediction of the observed data (Schemper and Henderson, 2000; Henderson et al., 2002; Gerds and Schumacher, 2006), while the second one involves discrimination measures and the assessment performance of the discrimination of the model between subjects who are more likely to experience the event shortly from subjects who are more likely to experience the event later (Heagerty and Zheng, 2005; Antolini et al., 2005; Pencina et al., 2008). Effect sizes for the association parameters in univariate joint models tend to be exaggerated, which is similar to the effect seen in standard regression models when covariates are omitted. A more insightful approach is to consider the area under the curve, or discrimination index, for both univariate and multivariate joint models in order to establish if the latter is able to better discriminate between individuals who had the event and those who did not. Such measures are scale-free and can be compared across biomarkers.

To follow, an assessment is made of how well a longitudinal biomarker is capable of discriminating between MCI patients who converted to AD and those who did not. Time-dependent areas under the Receiver Operating Characteristic (ROC) curves (AUCs) are found out in order to make an assessment of the predictive ability of the repeated measurements at different time points over the follow-up time for the time-to-event outcome. High AUCs indicate a strong capability of the biomarker in terms of discriminating between MCI patients who progress to AD within a short time frame from patients who do not.

The focus here is on a time interval of medical relevance and we consider the

dynamic nature of a longitudinal biomarker in terms of discriminating between individuals who experience the event within this time framework from the individuals who do not. As longitudinal measurements are collected over time, the predicted survival probabilities can be updated continuously. This feature of longitudinal data can help to discriminate between the two aforementioned sets of individuals. Let us assume that $\boldsymbol{Y_{ik}}(t) = \{\boldsymbol{y_{ik}}(s), 0 \leq s \leq t\}$ denotes the longitudinal measurements up to a time point t for subject i for the k^{th} biomarker. The interest lies in the chance of experiencing the event in the time frame $(t, t + \Delta t]$, within which an action can then be taken by a clinician to improve the chance of survival of the individual. Let $\pi_i(t, t + \Delta t|t)$ be the probability of an event for individual i in $(t, t + \Delta t]$ and consider the longitudinal history data $\boldsymbol{Y_{ik}}(t)$, and suppose individual i and i' are randomly chosen, where they each have measurements up to time t. The discriminating capability of the model can be assessed by the AUC as follows:

$$\text{AUC}(t, \Delta t) = \Pr\Big[\pi_i(t+\Delta t \mid t) < \pi_{i'}(t+\Delta t \mid t) \mid \{T_i^* \in (t, t+\Delta t]\} \cap \{T_{i'}^* > t+\Delta t\}\Big] \quad (6.5)$$

If the event occurred for the i^{th} individual, and did not occur for the i'^{th} individual within the time frame $(t, t + \Delta t]$, then if the model assigns higher survival probability for the individual who did not have the event then the model has successfully discriminated between individuals i and i'. Repeating this for all pairs gives the overall discrimination index.

6.6 Results

The demographic information of the study population is given in Table 6.2. Subjects who progressed to AD were followed up for a mean of 2.25 (1.79) years, while those who did not progress to AD were followed up for a mean of 4.53 (3.34) years. The mean baseline age was 74.75(7.28) for all participants. There were 139 female participants out of the 388 participants, and the mean of education length for all participants was 15.6(3.03) years. There were 131 participants who had event and had one or more APOE $\varepsilon 4$ alleles, whereas there were 80 participants who did not have event and had of one or more

	Progressed to AD ($n = 205$)	Did not progress to AD ($n = 183$)	Combined ($n = 388$)
Women	77(37.5%)	62(33.9%)	139(35.8%)
Age	74.53(7.04)	74.98(7.55)	74.75(7.28)
APOE4 present	131(63.9%)	80(43.71%)	211(54.38%)
Education (years)	15.83(2.85)	15.34(3.20)	15.60(3.03)
Time in study (years)	2.25(1.79)	4.53(3.34)	3.32(2.87)

Table 6.2: Baseline characteristics of ADNI-I participants with MCI. The data corresponding to women and APOE4 are given as number of participants (percentage), while the others are given as mean (SD).

APOE $\varepsilon 4$ alleles. Furthermore, 205 participants progressed to AD during the follow up period among the 388 MCI patients. The event rate is 52.8%.

Li et al. (2017) presented longitudinal trajectories for ADAS-Cog 13, ADAS-Cog 11, CDR-SB, and FDG-PET regardless of experiencing the event. Here, we aim to present the longitudinal profiles of these biomarkers for MCI patients who were converted to AD. Timescale is adjusted by taking away each patient's observed event time. Figure 6.1 represents the trajectories of ADAS-Cog 13 and ADAS-Cog 11, while Figure 6.2 represents the trajectories of CDR-SB and FDG-PET. It can be clearly seen that all measures deteriorated as AD progressed.

6.6.1 Results of the multivariate joint models

Tables 6.4 and 6.5 display the results of the joint models with time-to-dementia and multivariate longitudinal biomarkers and the corresponding score test. The multivariate joint models were fitted using the `mjoint` function in the `joineRML` package (Hickey et al., 2017). The null hypothesis is defined such that there is no association between time-to-dementia and the biomarker of interest in the model. The models in these tables are ranked arbitrarily. The column of p values in these tables shows that all γ values are found significant for the nine models (Models 4, 5, 6, 7, 9, 10, 15, 16, 18), one association parameter out of three is found significant for Model 14, and two association parameters out of three are found significant for the rest of the models. Whereas, according to the score test results, all the association parameters are needed in the

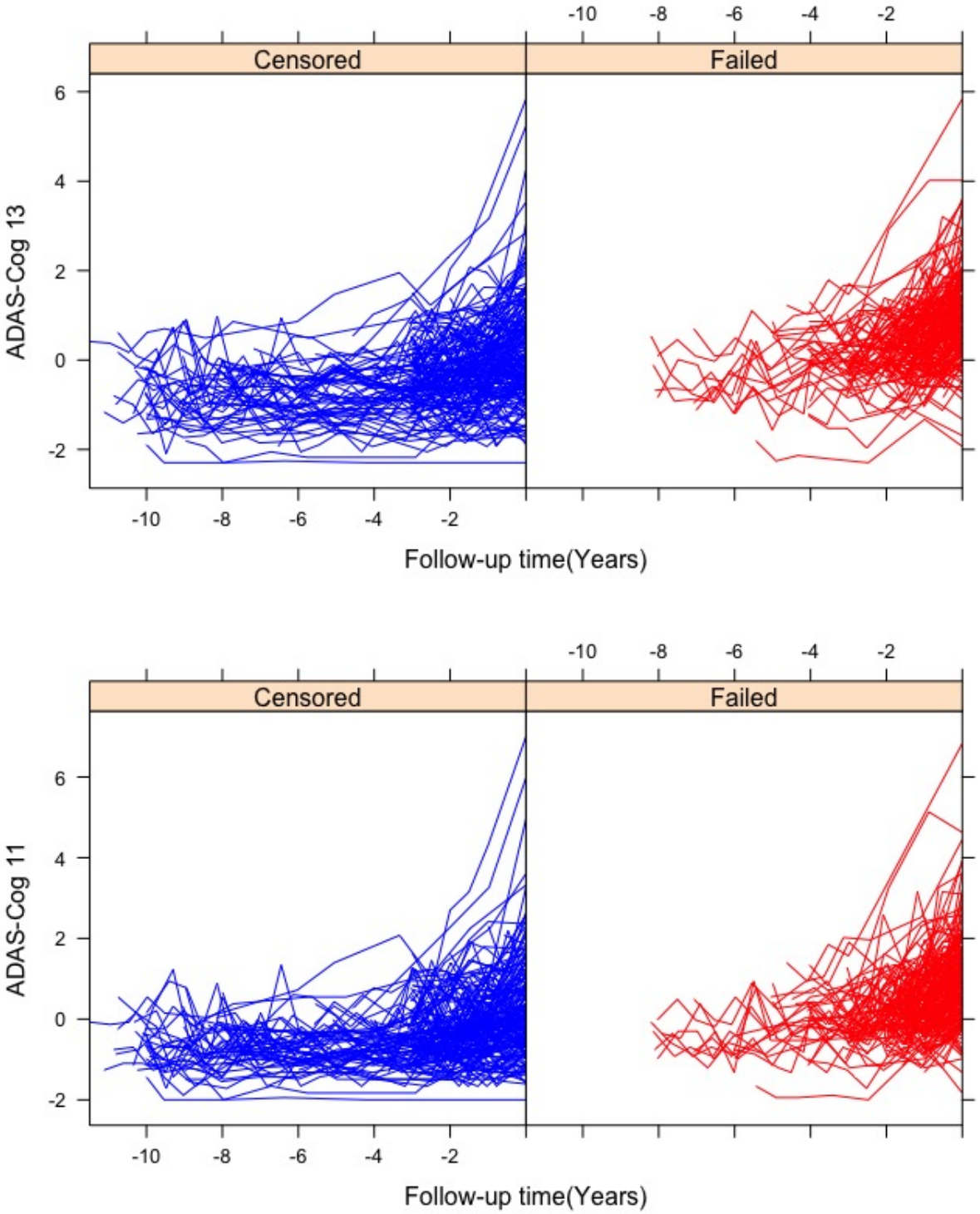

Figure 6.1: The longitudinal trajectories of each participants who are censored or not. The upper plot shows trajectories of ADAS-Cog 13, while the lower plot shows trajectories of ADAS-Cog 11. ADAS-Cog refers to Alzhemier's Disease Assessment Scale-Cognitive Subscale test.

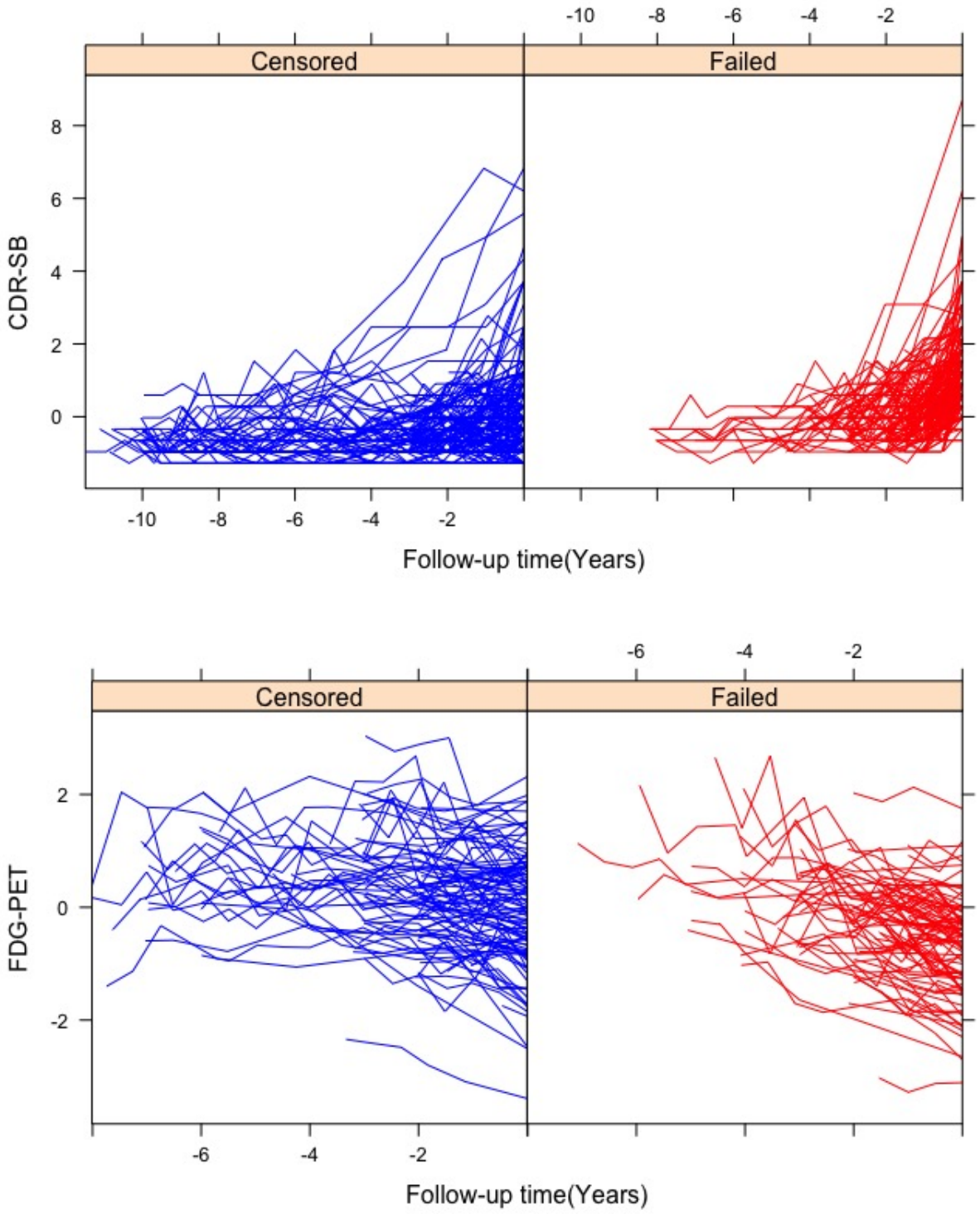

Figure 6.2: The longitudinal trajectories of each participants who are censored or not. The upper plot shows trajectories of CDR-SB, while the lower plot shows trajectories of FDG-PET. CDR-SB refers to Clinical Dementia Rating Sum of Boxes; whereas FDG-PET refers to the sum of mean glucose metabolism uptake in regions of angular, temporal, and posterior cingulate.

model for Models 1, 10 and 18, only one association parameter is needed in the model for Models 5 and 6, none of them is needed in the model for Models 2 and 11, and, although the γ parameter is at a reasonable level, two association parameters are needed in the rest of the 13 models.

6.6.2 Results of the univariate joint models

The model order in Table 6.6 is the same as in the corresponding table in Li et al. (2017) so that a proper comparison can be made between the results obtained via the packages `JM`, `joineR`, and `joineRML`. A column of Z values obtained via the `JM` package in Li et al. (2017) is also introduced here as Z* Value. This allows a comparison to be made between results obtained by those three packages and the score test. The `joineR` and `joineRML` packages results are either consistent or slightly different, where the difference can be considered negligible in terms of the association parameter. There are 23 significant biomarkers out of 33, as is the case in Li et al. (2017). The results obtained by the score test indicate that 24 longitudinal biomarkers (out of 33) are significant predictors of AD conversion. The longitudinal measures of EcogPtMem and MOCA were significant predictors, while RAVLT forgetting and AV45 were non-significant according to the results obtained via `joineR` and `joineRML`. On the other hand, according to the results obtained by the `JM` package, RAVLT forgetting and AV45 were found to be significant predictors, while the other two were non-significant. The longitudinal measures of EcogPtMem were non-significant according to the score test results, whilst the other three were significant predictors for time to AD conversion. The `JM` results are more different than the results obtained by the other two packages. This may be due to the fact that a slightly different joint model was fitted than the ones fitted in Li et al. (2017). Furthermore, there are some factors which may affect the performance of the joint model, such as the event rate and sample size. For instance, the repeated measurements of MOCA have 96 participants with a 31% event rate in our model, whereas the sample size was 77 with a 17% event rate in Li et al. (2017).

6.6.3 Results for score test

For more complicated scenarios, the score test is not as powerful as in univariate case studied in the previous chapter. The score test results and the results after fitting the MVJM are consistent only for 6 models (Models 3, 10, 17, 18, 19, 20) in Tables 6.4 and 6.5. That is, when we tested if a particular biomarker should be in the model before fitting the model, and when we actually fit the model with that particular biomarker and looked at the significance of the γ parameter, the score test is not a reliable tool for these scenarios.

As for the score test results, one can roughly assess the power of the test when γ is a particular value by looking at the empirical results obtained via the simulation studies in Chapter 4, Section 4.4. Figure 4.2 depicts the power of the score test based on three different scenarios and with 500 subjects. Table 6.3 indicates the power of the test when γ is -0.98, 0.31, 0.52, and 0.57 for the longitudinal measures of MOCA, RAVLT forgetting, AV45-PET, and EcogPtMem, respectively. The reason for choosing these four biomarkers is that there are inconsistencies between Z and Z^* values in Table 6.6. While the `joineRML` function finds the association parameters non-significant for RAVLT forgetting and AV45-PET, and significant for MOCA and EcogPtMem, the `JM` function finds otherwise. As the power is quite sensitive to the size of the random effect variance-covariance matrix, we only put Scenarios I and II to be close to the ADNI data, in which the random effect variance is much different than that represented by these scenarios. However, as the sample sizes of the biomarkers are much smaller than the sample size in the simulation studies, the power is affected negatively. According to the table, the power of the score test results based on bootstrap variance is higher than that based on the theoretical variance. As such, if the sample size is increased, the test could be more reliable. Despite the low power of the test with those parameters, the results in Table 6.6 are convincing and do not have much inconsistency among the joint modelling results obtained by the different packages. For example, as the closest biomarker to Scenarios I and II is RAVLT forgetting from Table 6.1 in terms of

	Biomarker	γ	$Power_{V_1}$	$Power_{boot}$	Event Rate	Sample Size
Scenario I	MOCA	-0.98	79	90	31.9	96
	RAVLT.forgetting	0.31	23	33	52.8	388
	AV45-PET	0.52	51	74	31.4	70
	EcogPtMem	0.57	40	71	30.2	96
Scenario II	MOCA	-0.98	17	30	31.9	96
	RAVLT.forgetting	0.31	9	18	52.8	388
	AV45-PET	0.52	11	23	31.4	70
	EcogPtMem	0.57	6	18	30.2	96

Table 6.3: Indication of the power of the score test in percentage for a random intercept and slope model with a representative small sample of the simulation results for Scenario I and II: Scenario I, 70% event rate; Scenario II, 25% event rate. Longitudinal and survival data are linked through the association parameter γ, with independence at $\gamma = 0$.

event rate and sample size, we looked at the power of the score test to assess it from Table 6.1 when we have these kinds of scenarios. The power of the test based on theoretical variance when γ is 0.31 (RAVLT forgetting) for Scenario I is 23%, while for Scenario II, it is 9%. As such, one can say that the power of the test for this scenario (i.e. RAVLT forgetting) is between 9% and 23%. While `JM` finds that this magnitude of association parameter estimate significant, `joineRML` finds that there is not enough evidence to be able to assess its significance, and the score test finds that it is significant with low power.

Model	Variable*	Participants	Observations	Events (%)	Joint Model (joineRML) γ	γ Value	Z Value	P Value	Score	P_{score}
Model 1	ADAS-Cog 13	388	2073	205(52.8%)	γ_{ADAS13}	0.2570	1.3328	0.1826	-4.031	0.0001*
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.6680	-2.7351	0.0062*	3.597	0.0003*
	FAQ				γ_{FAQ}	0.6228	8.6145	0.0001*	2.092	0.036*
Model 2	ADAS-Cog 13	362	1389	187(51.7%)	γ_{ADAS13}	0.8442	4.1768	0.0001*	0.080	0.936
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.2015	-0.8613	0.3891	1.565	0.118
	MidTemp				γ_{Mid}	-0.2277	-2.2861	0.0222*	-0.461	0.645
Model 3	ADAS-Cog 13	388	2094	205(52.8%)	γ_{ADAS13}	0.5167	2.5294	0.0114*	2.933	0.003*
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.0509	-0.1669	0.8674	-0.519	0.604
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-1.0014	-2.9672	0.0030*	-3.212	0.001*
Model 4	ADAS-Cog 13	362	1390	187(51.7%)	γ_{ADAS13}	0.6312	7.3626	0.0001*	-0.705	0.481
	FAQ				γ_{FAQ}	0.5100	6.7868	0.0001*	2.799	0.005*
	MidTemp				γ_{Mid}	-0.2139	-2.2104	0.0271*	3.341	0.001*
Model 5	ADAS-Cog 13	388	2073	205(52.8%)	γ_{ADAS13}	0.4980	3.8026	0.0001*	-1.228	0.219
	FAQ				γ_{FAQ}	0.5300	7.3731	0.0001*	1.407	0.159
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-0.6615	-2.3569	0.0184*	2.464	0.014*
Model 6	ADAS-Cog 13	362	1389	187(51.7%)	γ_{ADAS13}	0.7123	4.8317	0.0001*	2.715	0.007*
	MidTemp				γ_{Mid}	-0.2887	-2.7808	0.0054*	-1.132	0.257
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-0.6957	-2.9200	0.0035*	0.726	0.468
Model 7	FAQ	365	1394	188(51.5%)	γ_{FAQ}	0.4778	5.5850	0.0001*	4.316	0.00002*
	MidTemp				γ_{Mid}	-0.4236	-4.0236	0.0001*	-3.172	0.002*
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-1.0294	-5.1144	0.0001*	-1.610	0.107
Model 8	RAVLT.learning	388	2098	205(52.8%)	$\gamma_{RAVLT.learn}$	-0.4937	-1.4389	0.1502	3.797	0.0001*
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.7541	-3.6124	0.0003*	0.799	0.424
	FAQ				γ_{FAQ}	0.6270	9.1002	0.0001*	3.206	0.001*
Model 9	MidTemp	365	1394	188(51.5%)	γ_{Mid}	-0.2622	-2.5744	0.0100*	-1.012	0.311
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-0.9464	-7.0948	0.0001*	-4.077	0.00005*
	FAQ				γ_{FAQ}	0.5977	8.2958	0.0001*	4.181	0.00003*
Model 10	RAVLT.immediate	365	1401	188(51.5%)	$\gamma_{RAVLT.imm}$	-0.5302	-2.3837	0.0171*	-4.132	0.00004*
	MidTemp				γ_{Mid}	-0.3706	-3.5057	0.0005*	-3.988	0.0001*
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-0.8713	-2.8953	0.0038*	-2.463	0.014*

Table 6.4: The MVJM results for prediction of risk of AD progression. Fixed effects are baseline age, the presence of APOEε4 in the longitudinal submodel, gender, baseline age, years of education, and the presence of APOEε4 in the survival submodel. *Models are ranked arbitrarily.

Model	Variable*	Participants	Observations	Events (%)	Joint Model (joineRML) γ	γ Value	Z Value	P Value	Score	P_{score}
Model 11	ADAS-Cog 13	363	1399	188 (51.8%))	γ_{ADAS13}	0.7117	3.3799	0.0007*	-0.623	0.534
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.3822	-1.6271	0.1037	1.752	0.080
	Hippocampus				γ_{Hip}	-0.2475	-2.1780	0.0294*	0.106	0.915
Model 12	ADAS-Cog 13	363	1400	188 (51.8%)	γ_{ADAS13}	0.6622	7.9864	0.0001*	-0.877	0.381
	FAQ				γ_{FAQ}	0.4884	5.9646	0.0001*	2.003	0.045*
	Hippocampus				γ_{Hip}	-0.1465	-1.2932	0.1959	5.811	0*
Model 13	ADAS-Cog 13	362	1392	187 (51.7%)	γ_{ADAS13}	0.8057	8.2522	0.0001*	2.824	0.005*
	MidTemp				γ_{Mid}	-0.2327	-2.2681	0.0233*	1.523	0.128
	Hippocampus				γ_{Hip}	-0.1573	-1.3192	0.1871	3.555	0.0004*
Model 14	ADAS-Cog 13	363	1399	188(51.8%)	γ_{ADAS13}	0.8119	5.4011	0.0001*	2.982	0.003*
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-0.4920	-1.7807	0.0750	3.037	0.002*
	Hippocampus				γ_{Hip}	-0.1777	-1.5427	0.1229	-0.185	0.853
Model 15	FAQ	366	1404	189 (51.6%)	γ_{FAQ}	0.5754	7.4173	0.0001*	2.999	0.003*
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-1.0037	-7.9765	0.0001*	-4.858	0.00000*
	Hippocampus				γ_{Hip}	-0.2745	-2.4694	0.0135*	0.867	0.386
Model 16	MidTemp	365	1396	188 (51.5%)	γ_{Mid}	-0.2702	-2.7010	0.0069*	-2.088	0.037*
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.9818	-7.2892	0.0001*	-5.977	0*
	Hippocampus				γ_{Hip}	-0.3168	-2.9400	0.0033*	-1.093	0.274
Model 17	RAVLT.learning	366	1411	189 (51.6%)	$\gamma_{RAVLT.learn}$	-0.4295	-1.1746	0.2401	0.636	0.525
	RAVLT.immediate				$\gamma_{RAVLT.imm}$	-0.8543	-3.6103	0.0003*	-3.710	0.0002*
	Hippocampus				γ_{Hip}	-0.3399	-2.8553	0.0043*	-2.416	0.016*
Model 18	Hippocampus	365	1397	188 (51.5%)	γ_{Hip}	-0.2409	-2.2886	0.0221*	3.273	0.001*
	MidTemp				γ_{Mid}	-0.4326	-3.9835	0.0001*	-2.944	0.003*
	FAQ				γ_{FAQ}	0.6506	10.2153	0.0001*	3.907	0.0001*
Model 19	Hippocampus	366	1404	189 (51.6%)	γ_{Hip}	-0.1867	-1.6615	0.0966	0.261	0.794
	RAVLT.learning				$\gamma_{RAVLT.learn}$	-1.1068	-5.0580	0.0001*	2.597	0.009*
	FAQ				γ_{FAQ}	0.5088	5.3926	0.0001*	3.939	0.0001*
Model 20	RAVLT.learning	365	1396	188(51.5%)	$\gamma_{RAVLT.learn}$	-1.0791	-4.9784	0.0001*	-3.794	0.0001*
	MidTemp				γ_{Mid}	-0.5090	-4.5180	0.0001*	-3.862	0.0001*
	Hippocampus				γ_{Hip}	-0.1566	-1.1778	0.2389	-1.263	0.207

Table 6.5: The MVJM results for prediction of risk of AD progression for the second set of 10 models.

				Joint Model (JoineR)		Joint Model (joineRML)			Score Statistics				
Variable*	Participants	Observations	Events (%)	γ	HR	γ	HR	Z Value	$Score_{V1}$	(P_{V1})	$Score_{boot}$	(P_{boot})	Z* Value
ADAS-Cog 13	388	2107	205(52.8%)	0.92	2.50	0.92	2.51	14.38	9.055	0	9.329	0	9.33
ADAS-Cog 11	388	2126	205(52.8%)	0.75	2.12	0.76	2.12	13.24	7.959	0	8.411	0	8.40
RAVLT.immediate	388	2119	205(52.8%)	-1.28	3.59	-1.30	3.59	-14.22	-8.614	0	-10.380	0	-8.32
FAQ	388	2132	205(52.8%)	0.92	2.50	0.92	2.51	14.22	6.910	0	6.600	0	7.38
MidTemp	365	1410	188(51.5%)	-0.70	2.01	-0.71	2.02	-9.52	-6.804	0	-6.907	0	-7.31
RAVLT.learning	388	2119	205(52.8%)	-1.45	4.29	-1.45	4.28	-8.48	-7.528	0	-8.645	0	-6.45
MMSE	388	2128	205(52.8%)	-0.68	1.97	-0.69	1.99	-18.28	-6.411	0	-5.996	0	-6.43
Hippocampus	366	1420	189(51.6%)	-0.69	2.00	-0.70	2.01	-8.69	-6.401	0	-6.505	0	-6.30
CDR-SB	388	2134	205(52.8%)	0.73	2.08	0.74	2.09	19.73	3.296	0.001	3.620	0.0003	6.30
FDG-PET	227	9892	107 (47.1%)	-0.78	2.19	-0.78	2.18	-8.16	-5.596	0.00000	-5.799	0	-6.08
Entorhinal	365	1410	188 (51.5%)	-0.71	2.03	-0.71	2.04	-7.27	-6.329	0	-7.004	0	-6.00
Fusiform	365	1410	188(51.5%)	-0.56	1.76	-0.57	1.76	-6.82	-4.998	0.00000	-6.178	0	-5.33
EcogSPTotal	97	353	31 (32%)	1.31	3.70	1.32	3.73	5.10	3.212	0.001	2.568	0.010	3.24
WholeBrain	387	1779	204 (52.7%)	-0.42	1.52	-0.42	1.52	-4.51	-3.217	0.001	-3.431	0.001	-3.21
EcogSPMem	98	354	31(31.6%)	1.52	4.58	1.53	4.60	5.73	3.972	0.0001	3.394	0.001	3.14
EcogSPLang	98	355	31(31.6%)	0.89	2.44	0.89	2.44	3.54	2.349	0.019	2.686	0.007	2.96
EcogSPDivatt	97	327	31(32%)	1.17	3.22	1.17	3.22	4.02	3.419	0.001	2.061	0.039	2.93
EcogSPOrgan	94	337	28(29.8%)	1.20	3.31	1.20	3.31	4.30	3.539	0.0004	1.855	0.064	2.85
EcogSPPlan	97	347	31(32%)	1.19	3.27	1.19	3.28	4.60	2.879	0.004	2.135	0.033	2.66
RAVLT.forgetting	388	2113	205(52.8%)	0.31	1.36	0.31	1.36	1.65	2.133	0.033	2.090	0.037	2.65
EcogSPVisspat	94	339	29(30.9%)	1.05	2.87	1.05	2.87	4.53	2.916	0.004	2.400	0.016	2.49
Ventricles	386	1761	203(52.6%)	0.23	1.26	0.23	1.26	3.45	2.495	0.013	2.386	0.017	2.37
AV45-PET	70	134	22(31.4%)	0.52	1.67	0.52	1.67	1.71	2.341	0.019	1.994	0.046	2.01
EcogPtOrgan	95	353	29(30.2%)	0.27	1.32	0.28	1.32	1.05	1.137	0.255	1.115	0.265	1.95
EcogPtMem	96	359	29(30.2%)	0.57	1.77	0.57	1.77	2.03	1.739	0.082	2.028	0.043	1.92
EcogPtTotal	96	359	29(30.2%)	0.34	1.41	0.34	1.41	1.34	0.705	0.481	0.754	0.451	1.66
EcogPtDivatt	96	357	29(30.2%)	0.01	1.01	0.01	1.01	0.02	0.182	0.856	0.204	0.838	1.36
MOCA	94	344	30(31.99%)	-0.98	2.68	-0.99	2.69	-4.77	-3.678	0.0002	-3.105	0.002	-1.34
PIB-PET	61	125	28(45.9%)	0.31	1.36	0.32	1.38	1.42	-0.630	0.529	-0.560	0.576	1.31
EcogPtLang	96	358	29(30.2%)	0.11	1.11	0.11	1.11	0.40	0.199	0.843	0.201	0.841	1.08
EcogPtPlan	95	358	28(29.5%)	0.32	1.38	0.32	1.38	1.00	0.425	0.671	0.420	0.675	0.99
EcogPtVisspat	96	353	29(30.2%)	0.40	1.50	0.41	1.50	1.57	1.360	0.174	1.450	0.147	0.93
ICV	387	1803	204(52.7%)	0.05	1.05	0.05	1.05	0.50	0.051	0.959	0.061	0.951	0.05

Table 6.6: The univariate JM results for prediction of risk of AD progression. Fixed effects are baseline age, the presence of APOEε4 in the longitudinal submodel, gender, baseline age, years of education, and the presence of APOEε4 in the survival submodel. *Models are ranked based on the absolute value of the Z* Value for the association parameter.

6.6.4 Results of discrimination for multivariate longitudinal biomarkers and predictive accuracy

In this subsection, we illustrate the calculation of dynamic predictions from a trivariate joint model fitted to the ADNI dataset for the longitudinal outcomes ADAS-Cog 13, FAQ, and MidTemp. The biomarkers were chosen based on their significance within the univariate models (see Table 6.6). Due to the expected high correlation, the most significant univariate biomarker was selected from each of the three domains. Model 4 from Table 6.4 fitted with the most significant biomarker from each domain. The aim here is to calculate predictions of the conditional survival probabilities for Patient 1300 due to the sufficient number of measurements that they have. The dynamic survival probabilities as time progresses and the longitudinal trajectories of these biomarkers can be seen in Figure 6.3. Figure 6.4 depicts the calculation of the conditional expected longitudinal values for Patient 1300 from the last observation time, given their longitudinal history data. These figures present the survival probabilities predictions computed starting from the time point of the last measurements, up to the end of follow-up of the original dataset. It can be seen from Figure 6.4 that while the patient's cognitive function deteriorates due to the increasing ADAS-Cog 13, his functional and behavioural assessment, and neuroimaging function are getting better due to the increasing FAQ and MidTemp measurements. However, his survival probability decreases after 8 years, probably due to the increase of the overall hazard.

An investigation of what gains can be made by fitting the multivariate joint model over separately fitted univariate joint models was subsequently conducted. At this point, AUCs provide insight into choosing the optimum model. Model 4 from Table 6.6 was considered once again due to the significance within univariate models from each domain. First, the univariate joint models were fitted with each longitudinal biomarker separately, before subsequently fitting each possible bivariate model, and lastly, a trivariate model. ADAS-Cog 13 was the first biomarker for the univariate joint model, followed by FAQ, and then MidTemp. Pairs of the biomarkers for the bivariate joint models are

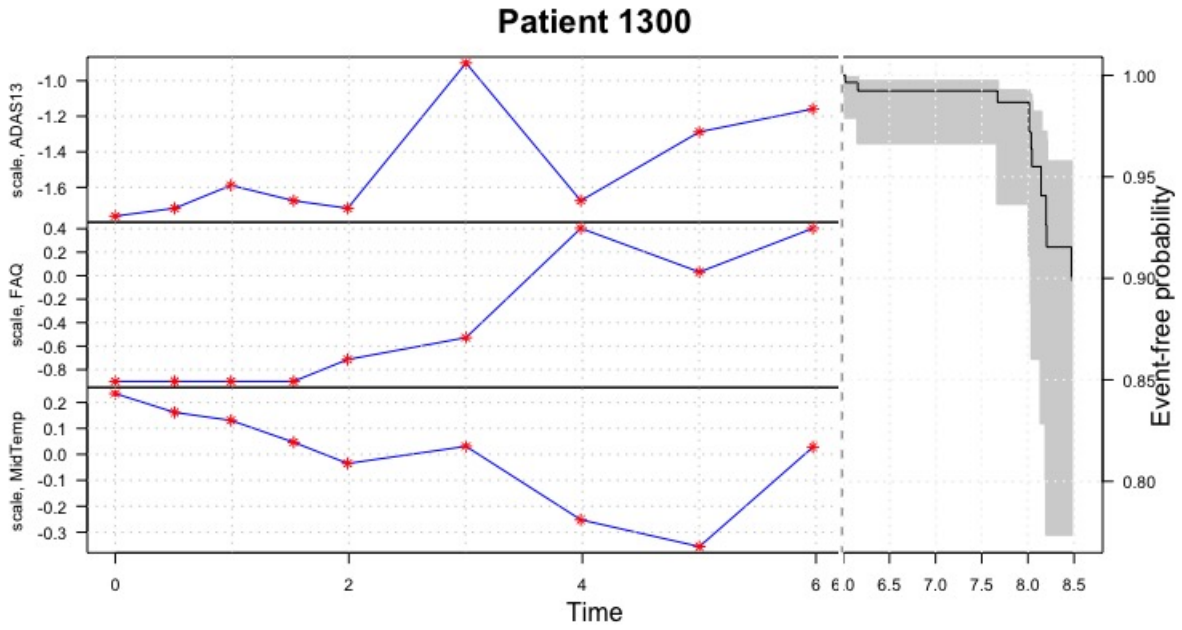

Figure 6.3: The conditional time-to-event distribution for a new subject from the last observation time, given their longitudinal history data.

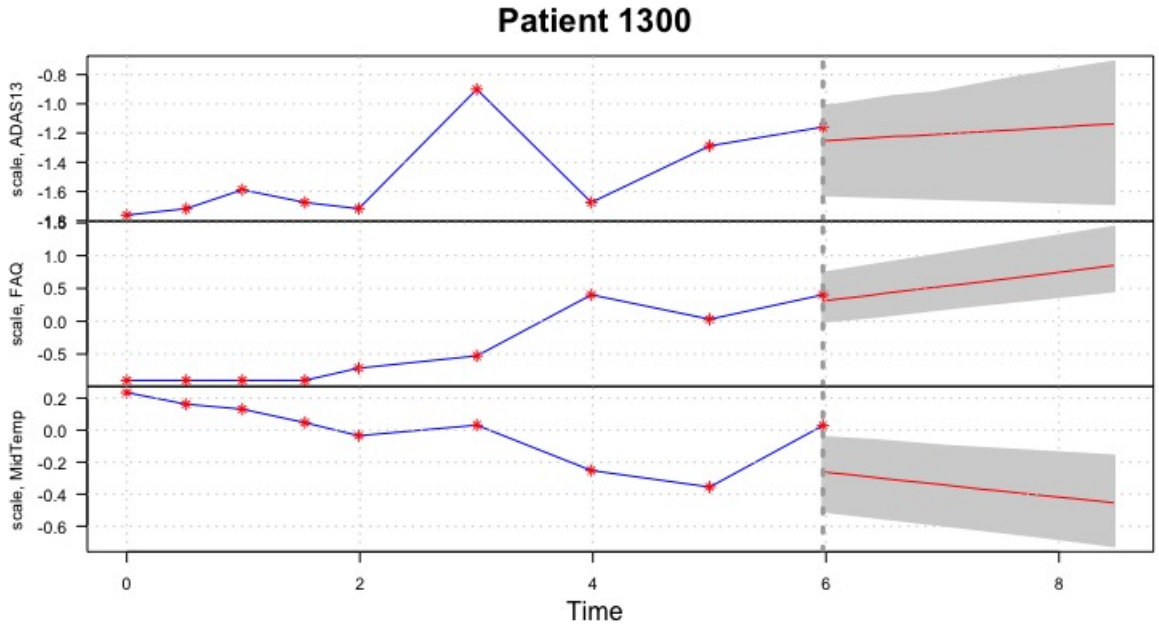

Figure 6.4: The conditional expected longitudinal values for Patient 1300 from the last observation time, given their longitudinal history data.

ADAS-Cog 13 - FAQ, ADAS-Cog 13 - MidTemp, and lastly FAQ - MidTemp. Figure 6.5 shows boxplots for AUCs for each model from 1000 Monte Carlo samples. We focus on time point $t = 1$ year and we are interested in the event occuring within a $\Delta t = 0.5$ year, 6 months interval. The figure indicates that the AUCs of the bivariate model of ADAS-Cog 13 - FAQ and the trivariate model are almost the same. Furthermore, there is not much to gain from adding the biomarker MidTemp. The longitudinal biomarkers of ADAS-Cog 13 and FAQ are good representatives for this dataset. The AUC of univariate model fits for each biomarker is almost equal to the gain of bivariate model fits for pairs ADAS-Cog 13 - MidTemp, and FAQ - MidTemp. The AUC results in the boxplot are presented as follows: the posterior means, with 2.5th and 97.5th percentiles in parentheses, for the six models (in the same order as the boxplot) are 0.78 (0.73 - 0.80), 0.77 (0.74 - 0.79), 0.71 (0.68 - 0.73), 0.82 (0.80 - 0.84), 0.77 (0.73 - 0.81) and 0.82 (0.79 - 0.85). These are based on 1000 simulations in each case.

6.7 Discussion

In this chapter, our aim was to present an extension of the univariate joint modelling framework from the longitudinal perspective, with multiple longitudinal outcomes and single event times data, and to gauge the extent to which a multivariate joint model may offer improved performance over a univariate joint model. The proposed framework considers the correlation between longitudinal outcomes and incorporates them with the event times. The principal advantage of this approach over fitting multiple separate univariate joint models is the enhancement of the predictive ability of the joint model, and the identification of the optimal combination of repeated outcomes in the determination of the hazard of the event.

The ADNI dataset, which contains 33 longitudinal biomarkers, was analysed regarding clinical and biological information of MCI patients from three domains, where the time-to-event outcome is the time for the conversion of MCI to AD. Our analysis includes joint models with multiple longitudinal biomark-

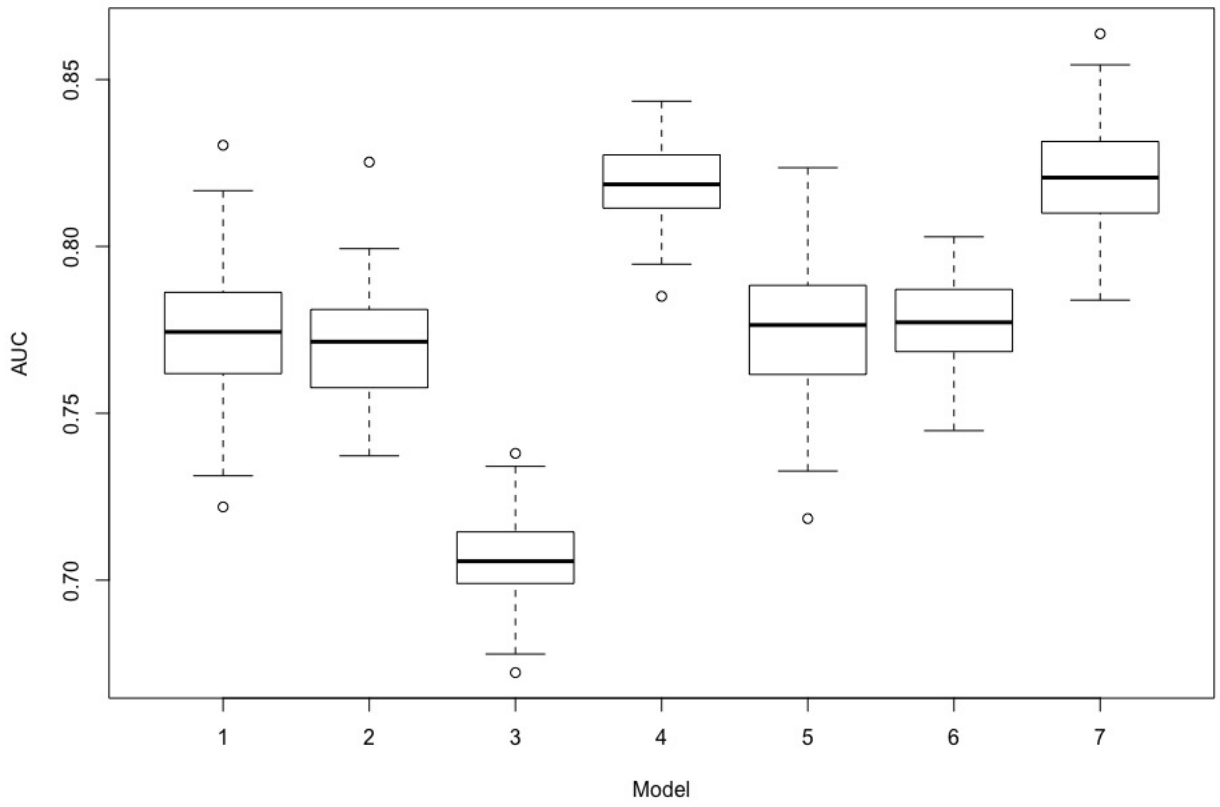

Figure 6.5: Boxplots of AUCs for all possible models including three biomarkers: ADAS-Cog 13, FAQ, and MidTemp. The order is univariate models for ADAS-Cog 13, FAQ, and MidTemp; bivariate models for pairs ADAS-Cog 13 - FAQ, ADAS-Cog 13 - MidTemp, and FAQ - MidTemp; and lastly, the trivariate model.

ers and event times, as well as separate joint models for each marker. In addition to these, a score test analysis was conducted, serving as a prognostic tool for the identification of the association between longitudinal and survival outcomes for both univariate and multivariate joint models. The results show how a biomarker is useful for the prediction of the hazard for the conversion of MCI to AD with and without some other biomarkers. In general, consideration of correlation between biomarkers may substantially improve the ability of prediction of the optimum treatment for the risk of conversion to AD in MCI patients. Furthermore, while the score test is a useful tool for the identification of appropriate longitudinal biomarkers in the hazard of conversion to AD for separate fitted joint models, it may not be as useful as this simple case for a combination of a couple of biomarkers and time-to-dementia. This is most likely explained by the fact that factors affecting the power of the test, such as the number of patients, event rate, longitudinal data structure (how balanced the data is), and correlation between biomarkers, are not good enough to result in a reasonable power level. Another explanation might be that the domains where each marker is from can affect the hazard of the conversion to AD. The domains may have different effect on the prediction of the hazard of the event. All trio combinations of the strongest predictors specified by Li et al. (2017) were used. The analyses in this chapter comprehensively and systematically evaluates the predictive ability of the combinations of markers for AD conversion under the MVJM framework.

No effort was made here to optimally choose biomarkers, as simply the most significant univariate predictor within each of the three domains was selected. Furthermore, while the functional form of the random effects in the longitudinal sub-models was an intercept-and-slope model throughout; in practice, this could be tailored to suit each individual biomarker. Similarly, the choice of association structure was a blanket assumption of proportionality, which may not be the optimal choice for each, or indeed, any of the biomarkers considered. In large datasets with many potential biomarkers, model selection considerations such as these are an open research area. Automation of such a process,

similar to that of stepwise selection, would be an asset to clinicians faced with such datasets in their area of work. Furthermore, the biomarkers may interact with one another, leading to potential interactions between the random effects from each longitudinal model within the survival model, adding another layer of complexity to the model selection problem. Currently, there is no software or methodology to combat such issues, and as such, this an avenue for future research.

Here, the focus was on looking at the AUCs at single time-point and time-interval, as this information can be combined over multiple intervals into what is termed a dynamic discrimination index, which, for a fixed length of time interval, can be used to weight the AUCs. The weighting is of potential importance, since as studies progress more, events and heavier censoring lead to less available data, and, hence, less reliable discrimination.

The areas under the curve calculated here may be over-estimates, since they are in-sample predictions. A cross-validation approach using training and test partitions of the data, K-fold cross-validation as in Li et al. (2017), or bootstrapping would guard against this, and is the focus of research for a publication of these initial findings based on the full sample. However, since the whole sample for each model is used here, one can anticipate that any overestimation will be similar in each case. A further comparison with the performance of the Cox proportional hazards survival models using the baseline measures of one, or more, biomarkers would help provide a fuller picture as to what gains can be made in moving from a separate analysis to a univariate joint model, and then further to a multivariate joint model.

The main disadvantage of this model is that MVJM is computationally challenging due to the increased dimensionality of the random effects component. We limit our application to trivariate data because of this reason. Application of this data for four or more longitudinal biomarkers methods for data dimension reduction can be an extension to the work of this chapter, as well as

investigating the power of the score test for different number of longitudinal biomarkers.

Chapter 7

Assessing robustness of univariate joint models for longitudinal and time-to-event data

7.1 Introduction

In Chapter 6, a joint model for time-to-event and multiple longitudinal data was defined, and an application of the ADNI study to this model as well as the univariate joint modelling was demonstrated for single event times data and single longitudinal data for each biomarker. Twenty multivariate joint models, consisting of all possible triples of the most significant six biomarkers found by Li et al. (2017), and 33 univariate joint models were obtained. The performance of the model varies subject to the characteristics of the dataset, such as sample size (since some biomarkers have a considerable amount of missing data), variance covariance matrix of the random effects parameters, noise, and number of measurement times per subject. Hence, it is intriguing to know how robust the parameter estimates of the model are under various parameterisations.

In joint models, the appropriate random effects parameter setting is important for obtaining more efficient inference procedures and for shedding light on the characteristics of the repeated measurements (Hsieh et al., 2006). Nonetheless, the model assumptions on random effects in joint models can affect the sen-

sitivity of inference. In this chapter, the issue of robustness of the univariate joint model is examined in the case of having various densities of longitudinal measurements to evaluate how parameter estimates vary in relation to their assumptions.

A typical example of a joint modelling setting is the AIDS Clinical Trials Group (ACTG) Protocol 175 (Hammer et al., 1996), in which repeated measurements of CD4 cell counts are believed to predict the time for progression to AIDS or the time of death. In this study, the CD4 cell counts are subject to assay errors and intra-subject variations.

Robustness in the joint modelling framework has been investigated in terms of various aspects via simulation studies, such as the robustness of model estimates to misspecification of errors, error structures themselves, and the magnitude of errors (Pantazis et al., 2005; Pantazis and Touloumi, 2007; Tang and Tang, 2015; Jaffa et al., 2011). Pantazis and Touloumi (2007) investigated the robustness of the model via simulation studies under a range of heavy tailed, skewed, and mixture distributions by fitting their proposed method, termed the joint multivariate random effects model (Pantazis et al., 2005). The authors concluded that the fixed effect parameter estimates in the longitudinal submodel seemed robust enough to misspecification, however their standard errors may be underestimated.

The methodology in univariate joint modelling does rely on assumptions about the random effects. Song et al. (2002b) and Rizopoulos et al. (2008) reported that joint models are quite robust to deviations from the assumption of normally distributed random effects. Diagnostic tools presented by Huang et al. (2009) can reveal adverse effects of random effect model misspecification in joint models of a primary endpoint and a longitudinal process.

The association structure used to link the event times via random effects relies on the repeated measurements. The random effects are generally assumed to

hold the normal distribution assumption. This assumption was investigated by Scharfstein et al. (1999) to determine whether or not inference is sensitive to the normality assumption. In this context, Song et al. (2002b) proposed a method to relax the distributional assumptions made for the random effects, and Tsiatis and Davidian (2001) proposed a method to make no parametric assumptions at all about the random effects. Nevertheless, the main finding from their approaches is that the parameter estimates are rather robust to misspecifying the random effects distribution. Rizopoulos et al. (2008) investigated the effect of misspecification of the random effects distribution in the shared parameter models, indicating that the effects of misspecified random effects diminish as the number of repeated measurements per subject increases. The authors derived the results theoretically, and considered in their simulation studies two cases for number of repeated measurements per subjects, m_i, namely the large-m_i case, where $max(m_i) = 15$ with 10 measurements per individual on average, and the small-m_i case, where $max(m_i) = 4$ with 2.5 measurements per individual on average, regardless of variance of the number of measurements per individual. A key question investigated here is whether or not the variance of the number of repeated measurements per individual has an impact on the robustness of the model, as well as the mean of the number of repeated measurements per individual, as only the latter was considered by Rizopoulos et al. (2008). To pursue this, simulation studies are conducted to identify the robustness of the model when the variance of the number of repeated measurements per subject changes. Since the univariate joint models dominate the current literature on joint modelling applications, the method that is applied in this chapter is the univariate joint modelling of longitudinal and time-to-event data, explained in Section 2.4.

In this chapter, robustness of the shared parameter models in the joint modelling framework is investigated. The method applied is briefly described in the following section. Section 7.3 details the simulation studies designed to investigate the sensitivity of the random effects assumptions in joint modelling, while Section 7.4 summarises the findings of the chapter and provides conclud-

ing remarks.

7.2 Method

In the framework of shared parameter models, joint models (JM), introduced by Wulfsohn and Tsiatis (1997), represent a useful parametrization method. This model employs a linear mixed model for a continuous longitudinal response and the corresponding hazard for the event at time t, as explained in detail in Section 2.4. This approach assumes that the (error-free) 'true' pattern of the longitudinal measurement, rather than the observed value, has impact on the hazard of an event. The 'true' pattern means the value of the predicted process at time t, as postulated by the underlying model. In this parametrization, the random effects are shared by the two processes of interest; longitudinal and survival. This means that the maximum likelihood estimate of the parameters has a contribution from both the longitudinal and the time-to-event processes.

7.3 Simulation study

7.3.1 Study design

The data generation method used throughout this thesis is deployed once more: using the `simjoint()` function in the `joineR` package (Philipson et al., 2017) (Section 2.7.1 can be referred to for a recap). The primary objective of this simulation study is to assess the robustness of the univariate joint modelling to the variability in the number of longitudinal measurements. The robustness of the model is assessed in terms of bias, mean square error (MSE) and coverage probability (CP).

We simulate data under three scenarios, and refer to them as Scenario I, II and III. The first scenario involves small mean and small variance of the number of measurements per subject, the second involves small mean and relatively large variance of the number of measurements per subject, while the third scenario involves large mean and large variance of the number of measurements per

subject. Figure 7.1 demonstrates indicative histograms for the number of measurement times per subject under Scenarios I, II and III. Approximately half of the subjects have one or two measurements for Scenarios I and II. The random

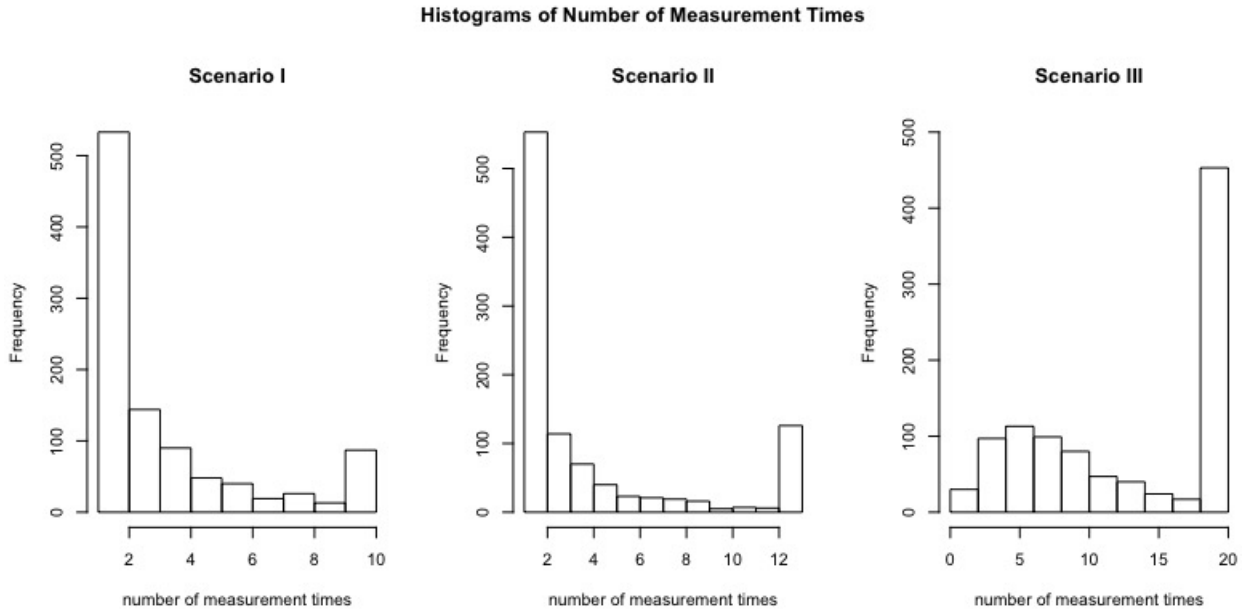

Figure 7.1: Indicative histograms based on a simulated dataset indicating the number of repeated measurement times per subjects for Scenarios I, II and III.

intercept and slope model is adopted, as it is the most commonly used model structure in practice, typically with a proportional association structure. This model, hence, is the basis of our simulation studies here. The chosen parameters to achieve the required scenarios are presented in Table 7.1. The mean of event rates of the simulated data vary across scenarios. The reason why these parameters were chosen was to see how sensitive the joint model is when there are lots of early dropouts from the study, which is very common in practice, and when most subjects remain until the end of study, which is less common, but can be encountered in practice, and to investigate the effect of variance of the number of measurement times per subject in conjunction with the mean of number of measurement times.

Each simulation is based on 100 repetitions with 100 bootstrap samples for each of three different sample sizes, $n = 250$, $n = 500$ and $n = 1000$, for each scenario. The balance of the simulated (synthetic) data varies across scenar-

	n	$E(m_i)$	$Var(m_i)$	Event Rate	$Max(m_i)$	θ_0	θ_1	λ_{cens}
Scenario I	250	3.23(0.166)	7.08(0.736)	57.34(3.001)	10	-2	0.5	exp(−2)
	500	3.23(0.117)	7.06(0.587)	57.28(2.423)	10	-2	0.5	exp(−2)
	1000	3.25(0.091)	7.15(0.428)	56.94(1.718)	10	-2	0.5	exp(−2)
Scenario II	250	3.99(0.267)	17.04(1.504)	71.83(2.798)	13	-2	1	exp(−3)
	500	3.91(0.183)	16.57(1.190)	72.46(1.961)	13	-2	1	exp(−3)
	1000	3.96(0.129)	16.86(0.851)	72.01(1.550)	13	-2	1	exp(−3)
Scenario III	250	13.24(0.424)	45.19(2.056)	49.83(3.449)	20	-7	0.2	exp(−4.5)
	500	13.37(0.306)	44.92(1.304)	49.65(2.290)	20	-7	0.2	exp(−4.5)
	1000	13.28(0.201)	45.15(0.943)	49.89(1.454)	20	-7	0.2	exp(−4.5)

Table 7.1: The parameters chosen for the implementation of scenarios in the simulation studies - mean and variance of the number of repeated measurements per subject with their SDs, and event rates experienced across 100 simulations.

ios, and therefore the computational time of the simulations varies depending on how balanced the longitudinal data is. While Scenario I took 35.9 hours to run the whole simulation, Scenario II and Scenario III took longer, with 175 hours and 255 hours, respectively. The reason may be that the first scenario has a small number of measurements per subject with less variability in the data structure, and as such, it does not take long to account for the available information.

7.3.2 Simulation results

To follow, each scenario is assessed separately, as the performance of the model is affected by the choice of parameters, which is particular to a given scenario. The fixed effects parameters, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$, are fixed in all of the scenarios as in Table 4.4. The association parameter, γ, and the error variances, σ_ε, can, on the other hand, vary to achieve the required scenarios.

Scenario I

The results of the parameter estimates are given in Table 7.2. All parameter estimates in the longitudinal submodel, excluding the time variable, are quite robust to misspecification. The coverage probability of the time variable decreases as the sample size increases. The bias appears to be fundamental, and

since variance decreases as sample size increases, one can be more confident of a biased result, which decreases the coverage probability.

More than half of the population of each simulation dropout very early from the study. Furthermore, the simulated data is quite unbalanced. Approximately 10% of the subjects have the maximum number of measurement times ($\max m_i$). Estimating the time coefficients is quite tricky in this scenario, as often only a few measurements are available for many individuals. This makes the model fragile.

Scenario II

The results of the parameter estimates under this scenario are given in Table 7.3. The most striking result in this scenario is that the bias of the association parameter estimate is considerably high in the negative direction. The magnitude of the association affects the parameter estimates (which are linked through the association parameter), such as the fixed effect parameter estimates in the time-to-event submodel. The parameter estimate for the time variable in the longitudinal submodel behaves in the same manner as in Scenario I, but to a greater extent, as CP is low even when $n = 250$. Furthermore, the parameter estimate of the variance for the random slopes, σ^2_{u1}, is also affected by the time variable. Surprisingly this did not occur under Scenario I. The CP of this is lower than those time-independent random effects parameter estimates, σ^2_{u0}. The data in this scenario is noisier than the data in Scenario I, and one can observe that the CP of noise is smaller compared to Scenario I. This is perhaps because the model variance owing to the random slopes was incorrectly captured through the general noise term. Another noteworthy point is that the performance of the model for the parameter estimates that exhibit bias deteriorates as the sample size increases again. This is due to the stronger certainty about a biased estimate. Furthermore, the ratio of the early dropout from the study may be a contributing factor to this. Owing to this, the model may have limited information for estimating the parameters. The mean parameter estimates are not noticeably different in terms of bias across sample sizes. However, as the

sample size increases, the MSE decreases, as well as the CPs.

Scenario III

Table 7.4 presents the parameter estimates results obtained from the joint analysis under Scenario III. This scenario represents the case in which the mean of the number of repeated measurements per subject is approximately 13, with high variability, as demonstrated in Table 7.1. The bias for all parameter estimates are at the nominal level.

The model performs very well. It can be said that as long as there is sufficient information to be able to estimate the parameters, the model based on an unbalanced dataset can perform well. Comparing directly Scenarios I and III, One can see that the model performed poorly for the time variable in Scenario I, where the early dropouts rate is almost half of the population in the simulated datasets, whereas all parameter estimates in Scenario III demonstrate good properties, except for γ when the sample size is large.

From the histograms depicted in Figure 7.1, it can be seen that approximately half of the subjects have one or two measurements under Scenarios I and II. This is the reason why the parameter estimates for the time variable in the longitudinal submodel in those scenarios are highly biased. Researchers should be encouraged to assess the distribution of the number of measurements when considering fitting a joint model that includes time in both the fixed and random components. This is not typically done, but is a recommendation of this thesis.

The histogram of the number of repeated measurements per subject under Scenario III shows that approximately half of the population is truncated at the maximum follow-up time. Since the mean of the longitudinal measurements per subject is high, the model finds sufficient information to be able to estimate the parameters in both the longitudinal and time-to-event submodels, along with the respective variance components. In addition, the peaks at the end

(owing to truncation) are typically censored individuals, who carry less information than those who have the event of interest.

When Scenarios I and II are compared, one can state that the smaller the variance of the number of repeated measurements per subject the study has, the more robust the joint model is. However, estimation of the time variable coefficient depends heavily on the magnitude of the number of repeated measurements per subject.

		$n = 250$			$n = 500$			$n = 1000$		
Variable	True Values	Bias	MSE	CP (95%)	Bias	MSE	CP (95%)	Bias	MSE	CP (95%)
Longitudinal										
Constant	7	-0.005	0.008	0.970	0.003	0.004	0.990	-0.001	0.003	0.930
Continuous	-1	-0.012	0.005	0.920	0.0003	0.003	0.900	-0.001	0.001	0.960
Binary	0.500	-0.002	0.013	0.970	-0.005	0.008	0.960	-0.001	0.004	0.940
Time	0.100	-0.107	0.018	0.710	-0.118	0.017	0.390	-0.129	0.018	0.060
Survival										
Continuous	0.600	0.001	0.016	0.920	-0.004	0.008	0.930	-0.009	0.003	0.940
Binary	-0.600	0.011	0.051	0.940	-0.016	0.022	0.980	0.024	0.012	0.940
Association										
γ	1	-0.002	0.005	0.970	-0.008	0.003	0.910	-0.013	0.002	0.910
Random Effects										
σ^2_{u0}	1	-0.013	0.006	0.970	-0.011	0.004	0.930	0.002	0.002	0.960
σ^2_{u1}	1	-0.023	0.013	0.910	-0.019	0.007	0.920	-0.016	0.003	0.940
Noise										
σ^2_{ε}	0.010	-0.0001	0.00000	0.960	0.00003	0.00000	0.930	0.00002	0.00000	0.950

Table 7.2: Simulation results of bias MSEs and coverage probabilities of joint model parameter estimates under Scenario I.

		$n = 250$			$n = 500$			$n = 1000$		
Variable	True Values	Bias	MSE	CP (95%)	Bias	MSE	CP (95%)	Bias	MSE	CP (95%)
Longitudinal										
Constant	7	-0.033	0.018	0.900	-0.033	0.008	0.920	-0.015	0.004	0.970
Continuous	-1	0.005	0.009	0.910	0.007	0.003	0.990	-0.0001	0.002	0.940
Binary	0.5	0.022	0.029	0.950	0.0004	0.013	0.940	-0.019	0.008	0.940
Time	0.1	0.249	0.072	0.240	0.212	0.051	0.240	0.224	0.054	0.080
Survival										
Continuous	0.6	-0.187	0.074	0.770	-0.201	0.056	0.620	-0.179	0.043	0.520
Binary	-0.6	0.154	0.182	0.910	0.194	0.093	0.920	0.219	0.085	0.760
Association										
γ	-3	1.053	1.181	0.090	1.041	1.122	0.040	1.023	1.063	0
Random Effects										
σ^2_{u0}	0.900	0.087	0.028	0.900	0.099	0.021	0.850	0.087	0.014	0.750
σ^2_{u1}	0.300	-0.057	0.007	0.660	-0.057	0.005	0.630	-0.061	0.005	0.430
Noise										
σ^2_{ε}	1	-0.033	0.004	0.830	-0.023	0.002	0.900	-0.024	0.001	0.870

Table 7.3: Simulation results of bias MSEs and coverage probabilities of joint model parameter estimates under Scenario II.

		$n = 250$			$n = 500$			$n = 1000$		
Variable	True Values	Bias	MSE	CP (95%)	Bias	MSE	CP (95%)	Bias	MSE	CP (95%)
Longitudinal										
Constant	7	-0.002	0.005	0.990	0.003	0.004	0.940	-0.006	0.002	0.950
Continuous	-1	0.007	0.004	0.970	0.007	0.003	0.900	-0.00002	0.001	0.970
Binary	0.500	0.006	0.012	1	0.005	0.009	0.960	-0.0003	0.004	0.940
Time	0.100	-0.010	0.004	0.970	0.007	0.002	0.910	-0.002	0.001	0.940
Survival										
Continuous	0.600	-0.035	0.013	0.960	-0.033	0.009	0.900	-0.015	0.003	0.960
Binary	-0.600	0.043	0.056	0.940	-0.007	0.031	0.900	0.019	0.010	0.980
Association										
γ	-1	0.034	0.006	0.960	0.039	0.004	0.920	0.041	0.005	0.880
Random Effects										
σ^2_{u0}	1	-0.009	0.008	0.940	-0.011	0.004	0.900	-0.001	0.002	0.970
σ^2_{u1}	1	-0.003	0.007	0.960	-0.001	0.004	0.930	0.004	0.002	0.950
Noise										
σ^2_ε	0.010	-0.00001	0.00000	0.940	0.00002	0.00000	0.920	-0.00002	0.00000	0.700

Table 7.4: Simulation results of bias MSEs and coverage probabilities of joint model parameter estimates under Scenario III.

7.4 Discussion

In this chapter, the robustness of random intercept and slope joint models was evaluated. The nature of the longitudinal trajectories over time clearly has an impact on the event time outcomes. The uncertainty of the random effects is pervasive in joint models. Rizopoulos et al. (2008) and Huang et al. (2009) have shown that as the number of repeated measurements per subject increases, the integrands converge to a normal density. Furthermore, the authors showed that the robustness of the model increases with the number of observations per subject. However, the potential question arising here is how the conclusions change when the balance or variability of the repeated longitudinal measurements per individual changes, since, clearly, not all individuals will have the mean number of measurements.

Simulation studies were conducted to investigate the effect of the length of the longitudinal trajectories, $E(m_i)$, and the homogeneity of these trajectories, $Var(m_i)$, in the shared parameter models in the joint modelling framework. The simulation studies raise two important issues: first of all, when there are a considerable number of early dropouts from the study, the performance of the model is directly affected, and the model becomes fragile; particularly in terms of the parameter estimate for the 'time' variable in the longitudinal submodel, and those including time-related measurements such as $\sigma^2_{U_1}$. The reason may be that the model cannot find enough information to be able to estimate these parameters unbiasedly. The second issue raised by these simulation studies is that the CP of the model decreases as the sample size increases. This may be due to the fact that as the sample size increases, the standard errors become smaller, which results in narrow confidence intervals and small CPs. It is worth noting that the bias is not a function of sample size, but a systematic failing of the model.

These results seem to be consistent with other research findings which show that as the number of repeated measurements per subject increases, the effect

of misspecification of the random effects distribution becomes minimal (Rizopoulos et al., 2008). A recommendation to researchers who deal with the joint model would be to investigate the mean and variance of the number of repeated measurements per subject as an exploratory data analysis tool, as it should help to understand how robust the joint model would be.

Chapter 8

Conclusions and Further Work

8.1 Introduction

In this thesis, a methodology and software have been developed for extending joint models of longitudinal and time-to-event data. Alongside this, the properties of joint models subject to heavily imbalanced data have been investigated, and a score test for determining how rich/complex a joint model needs to be has been developed. The methods have been assessed through simulation and application to a variety of clinical datasets. In the following section, a summary of the methodological developments described in previous chapters is presented. This chapter concludes the thesis as follows: Section 8.2 provides a summary of each chapter, as well as highlights the main original contributions, while Section 8.3 describes the limitations and possible extensions to the methodological frameworks of joint modelling described previously. Finally, the chapter is concluded by making final remarks in Section 8.4.

8.2 Summary of the thesis

The work presented in this thesis focuses on the methodological developments of and extensions to the joint modelling of longitudinal and time-to-event data. The findings from this study make several contributions to the current literature. In this section, a broad summary of the thesis is provided, conditions for the usage of the score test are explained, and the contributions corresponding

to each chapter are highlighted.

In Chapter 2, the foundations of the standard joint modelling proposed by Wulfsohn and Tsiatis (1997) and extended by Henderson et al. (2000) were described, as well as its two fundamental outcomes: longitudinal and time-to-event data. A few simulation studies were carried out to assess the model under a variety of scenarios, both with bootstrapping and Monte Carlo simulations. In addition, the behaviour of the model under a misspecified random effects model for all possible cases was investigated. The liver cirrhosis study was analysed, and the results were discussed. The assessment of the model through the first simulation was a necessary step to develop the code and build the extensions to the model in the subsequent chapters. Furthermore, such a study is important in terms of encouraging the use of joint models in the case of having repeated measurements, as well as event time data. This chapter also highlighted the superiority of the model over both the separate analysis approach, and the correctly specified random effects model. The simulation of the quadratic model required the development of a novel way to generate data.

In Chapter 3, the method and the software presented in Chapter 2 was developed to the multilevel platform, in which the joint model contains multilevel longitudinal and single time-to-event data. This method was examined through simulation studies, as well as the effects of misspecified random effects models. The work was motivated by a trial in a scleroderma lung study, in which the primary outcomes of the study were the forced vital capacity (FVC) measurements and death/failure times. The developed method employed the maximum likelihood estimation as the main estimation method used in the standard joint modelling framework. Taken together, these findings demonstrated a role for reducing the noise by considering the subject level and centre level heterogeneity. The model extended the scenarios, where clustering was implemented at the family level, or to further hierarchical terms.

In Chapter 4, a score test for association derived by Henderson et al. (2002)

was extended and implemented. Simulation studies were conducted to assess the power of the score test with the implementation of four types of random effects models under three different scenarios, and with applications made to the scleroderma lung study once more. The score test was used in conjunction with a proportional association structure. Two different variance estimators were considered.

Based on the simulation studies carried out for investigation of the power of the score test, a short guideline can be presented in order to when to use score test. The univariate score test is a very powerful tool in case of having strong association between longitudinal and survival outcomes regardless of positive or negative direction. This test is recommended to apply when the random effect model is time independent and when there are large number of subjects experienced the event. As the variability among subjects increases, the score test becomes more powerful and gives sensible results. If there are small number of subjects experienced the event, even with the simplest random effect model (Model A) and with a strong association, the score test is still not 100% reliable tool.

In Chapter 5, a considerable insight was gained with regard to the score test for association from a variety of cases, such as a score test with a separate association of each random effect component, a score test for multiple repeated longitudinal measurements, and a score test for multilevel framework in the joint modelling. These extensions were examined through simulation studies and an application of the SLS data. The score test is a powerful tool that can be used to reveal the need for the association between longitudinal and time-to-event data. Assessing the power of the score test under a variety of situations in this chapter and in Chapter 4 is important in order to gain an insight into the reliability of the test. Moreover, as trial protocols become more advanced, there is a need for prognostic tools that can help to identify the underlying complexity and the nature of the joint model that might be considered.

There are two important recommendations in order to decide using score test for complex joint models. First of all, the score test based on Martingale variance estimates are more reliable than the score test based on bootstrap variance estimates when there are sufficient number of subjects experienced the event with low subject variability. Bootstrap estimates are underestimated when the random effect model is time dependent (see Figure 5.1), they are overestimates when the random effect model is time independent (see Figure 5.2) in case of presence of multiple association parameters for each random effect component. Another recommendation for the clinicians can be that when multiple longitudinal profiles with sufficient event rates and strong associated longitudinal and survival data are present, the score test can be an applicable tool to decide whether which biomarker/s should be associated with the event of interest. This can help to make a critical decision which is able to change the direction of the treatment.

In Chapter 6, a joint model with multiple correlated longitudinal and a single survival outcomes was defined to demonstrate the usage of the optimum information in the case of having multiple consecutive repeated measurements. An application to the ADNI data was demonstrated to investigate the association between longitudinal biomarkers and time-to-dementia. Dynamic survival probabilities for an individual and discrimination capability of a biomarker were also assessed in order to investigate what gain can be achieved by fitting a multivariate joint model over fitting separate joint models for each biomarker. This is an important study that can be used to improve the predictive capability of the model, and lead to more informative inferences for the treatment of the Alzheimer disease with the incorporation of as much information as possible from the data.

In Chapter 7, the robustness of the univariate random intercept and slope joint model for three sample sizes was assessed via simulation studies in terms of how much variability in the repeated measurements should be. This study confirmed the technical findings of Rizopoulos et al. (2008), which had es-

tablished that the effect of random effects misspecification becomes minimal for certain parameters such as the number of repeated longitudinal measurements per subject. The main message was to consider the mean and variance of the number of longitudinal measurements in joint modelling dataset as an exploratory data analysis tool, in order to understand the strength of the predictive ability of the joint model.

8.3 Limitations and future work

Research into joint models continues to expand in a variety of aspects. One possible extension would be to extend the general multilevel framework described in Chapter 3 for more complex random effects model structures, such as random intercept and slope model at subject and centre level, and random intercept model at centre level, with intercept and slope model at subject level. Essentially, a bespoke association structure ought to be available for the end clinical user. Research into investigating the multilevel joint models with random intercept and slope model is in progress. The general multilevel framework can also be extended to incorporate frailty terms in the random effects model. The work conducted in Chapter 3 will be formally developed in the future as an R package to provide software to aid researchers analysing multilevel joint data.

Another avenue of progress for this project would be the development of flexible joint models with multiple longitudinal profiles and different forms of random effects models. There are numerous factors that affect the performance of the joint model in the univariate case. For instance, these factors could present data with different event rates, number of repeated measurements, levels of between subject heterogeneity, levels of association, random effects models, and sample sizes. When there are multiple longitudinal profiles, there is potential for interactions between profiles.

Allowing competing risks in the time-to-event outcome of the joint models

would be another useful extension. Competing risks are pervasive in clinical trials, in which patients are at risk of experiencing more than one event type, with the happening of an event type preventing the happening of the others (Putter et al., 2007). A common example is the death due to a couple of reasons, such as cancer and cardiovascular disease (Eloranta et al., 2012). While allowing competing risks in the joint models to be accounted for in a single study already exists in the literature (Williamson et al., 2008), joint models allowing the competing risks along with multilevel repeated measurements would be a major positive development.

Allowing parametric models would also be a useful extension of this work. Despite the fact that the Cox model is the most widely-used survival model (Cox, 1972), the interest in the parametric model has been increasing over the last decade (King et al., 2012) due to a number of benefits of the parametric approach in the time-to-event analysis. Crowther (2014) focused on the development of parametric models to analyse complex event time data, which includes the extensions to joint models of longitudinal and survival data. These extensions can be incorporated into a multilevel platform, as well as multivariate longitudinal profiles.

In Chapter 3, it was noted that a fundamental restriction usage of Gaussian quadrature approach in the EM algorithm for complex joint model structures is that the exponent has to be at least the size of the number of random effects. The weight of the abscissas has to be p^q (where p is the number of quadrature points, and q is the number of random effects components), which will be large once the random effects go large. As such, one important future research is to develop an alternative estimation method. The Monte Carlo EM algorithm (Hickey et al., 2018) considered for multivariate joint models, Laplace approximations considered for joint models (which are more computationally efficient than the Gaussian quadrature and Monte Carlo discussed by Rizopoulos et al. (2009) and Lin et al. (2008)), or the potentially quicker EM algorithm proposed by Meng and Van Dyk (1997) may be considered in order to accelerate the com-

putational time of the model.

8.4 Conclusion

In conclusion, the research presented in this thesis has examined joint models with longitudinal and time-to-event data from various aspects and through different means, including conducting a comparison between joint models and the separate analysis of the two components, considering an extension of the model to the multilevel platform, identifying association from a variety of aspects, describing the methodology of the joint model for multiple correlated longitudinal measurements, and investigating the robustness of joint models. This thesis is therefore a hybrid of simulation and application of real data. As the joint modelling is an optimal way of analysing data consisting of repeated measurements and event times, the areas of application are rather vast. Providing methodological developments as well as statistical software tools are thus of great value in enhancing the research area.

Appendices

Appendix A

Appendix

A.1 Calculation of the score and information

The derivatives of the expectation of the complete log-likelihood conditional on the observed data is set to zero in order to evaluate the parameter of estimates in M-step. Let call observed data D_0. The observed score and information of the current parameter estimates are necessary for Newton-Raphson algorithm. For example, let θ denote the current estimate of the parameter of interest, the Newton-Raphson algorithm is applied as follows:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \boldsymbol{I}_{\boldsymbol{\theta}}^{-1} \boldsymbol{S}_{\boldsymbol{\theta}} \tag{A.1}$$

where $\boldsymbol{I}_{\boldsymbol{\theta}}$ is the information and $\boldsymbol{S}_{\boldsymbol{\theta}}$ is the observed score.

The score of the parameter estimates is calculated as follows:

The conditional expectation of the complete log-likelihood as shown in (3.8) makes up of three densities. We will deal with this equation in three parts to make the calculation steps more understandable. The first density is regarding the longitudinal data, the second one is regarding the random effects and the last component is related to the survival part. Therefore, the first density only includes the parameter σ_ε^2 and the conditional expectation of its complete log-likelihood given the observed data is

$$E\left\{\log \prod_{h=1}^{H} \prod_{i=1}^{n_h} \prod_{j=1}^{J_i} f(y_{hij}|\boldsymbol{U_{hi}}, \sigma_\varepsilon^2)|D_0\right\} = -\frac{1}{2}\sum_{i=1}^{n_h} J_i \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n_h}\sum_{j=1}^{J_i} E(y_{hij} - \boldsymbol{x_{1i}}\boldsymbol{\beta_1} - \boldsymbol{D_{hi}}\boldsymbol{U_{hi}})^2$$

and the derivative of it set to zero, i.e.,

$$-\frac{1}{2\sigma_\varepsilon^2}\sum_{h=1}^{H}\sum_{i=1}^{n_h}J_i+\frac{1}{2\sigma_\varepsilon^4}\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{J_i}E(y_{hij}-\boldsymbol{x_{1i}}\boldsymbol{\beta_1}-\boldsymbol{D_{hi}}\boldsymbol{U_{hi}})^2=0$$

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h}J_i=\frac{1}{\sigma_\varepsilon^2}\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{J_i}E(y_{hij}-\boldsymbol{x_{1i}}\boldsymbol{\beta_1}-\boldsymbol{D_{hi}}\boldsymbol{U_{hi}})^2$$

$$\hat{\sigma_\varepsilon}^2=\sum_{h=1}^{H}\frac{\sum_{i=1}^{n_h}\sum_{j=1}^{J_i}E(y_{hij}-\boldsymbol{x_{1i}}\boldsymbol{\beta_1}-\boldsymbol{D_{hi}}\boldsymbol{U_{hi}})^2}{\sum_{i=1}^{n_h}J_i}$$

as shown in (3.11).

The second component of the complete log-likelihood contains $\boldsymbol{U}$ and Σ. The conditional expectation of this part of the complete log-likelihood is

$$E\left\{\log\prod_{h=1}^{H}\prod_{i=1}^{n_h}f(\boldsymbol{U_{hi}},|\boldsymbol{U},\Sigma)|D_0\right\}=-\frac{n}{2}\log|\Sigma|-\frac{1}{2}\sum_{i=1}^{n_h}E_i\{(\boldsymbol{U_{hi}}-\boldsymbol{U})'|\Sigma|^{-1}(\boldsymbol{U_{hi}}-\boldsymbol{U})\}$$

and the derivative of this term with respect to U is set to zero, i.e.,

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h}E_i\{(\boldsymbol{U_{hi}}-\boldsymbol{U})\}=0$$

$$\hat{\boldsymbol{U}}=\frac{\sum_{h=1}^{H}\sum_{i=1}^{n_h}E_i(\boldsymbol{U_{hi}})}{n}$$

as it can be given by equation (3.9). The derivative of this term with respect to Σ is set to zero and the maximum likelihood estimate of Σ is given in equation (3.10).

The third component of the complete log-likelihood is made up of the parameters $\boldsymbol{h_0}$, $\boldsymbol{\beta_2}$ and $\boldsymbol{\gamma}$.

$$E\left\{\log\prod_{h=1}^{H}\prod_{i=1}^{n_h}f(T_{hi},\delta_{hi}|\boldsymbol{U_{hi}},\boldsymbol{h_0},\boldsymbol{\gamma},\boldsymbol{\beta_2})|D_0\right\}=\sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\log\boldsymbol{h_0}(T_{hi})$$

$$+\sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}E_i\{\boldsymbol{x_{2i}}\boldsymbol{\beta_2}+\boldsymbol{\gamma}\boldsymbol{D_{hi}}\boldsymbol{U_{hi}}\}-\sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\int_0^{T_{hi}}h_o(s)E_i\{\ exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2}+\boldsymbol{\gamma}\boldsymbol{D_{hi}}\boldsymbol{U_{hi}})\}ds$$

Differentiating with respect to $h_0(s)$, we get

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h}\left\{\frac{\delta_{hi}I(T_{hi}=s)}{h_0(s)}-E_i\{\ exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2}+\boldsymbol{\gamma}\boldsymbol{D_{hi}}\boldsymbol{U_{hi}})\}Z_{hi}(s)\right\}\tag{A.2}$$

where $Z_{hi}(s)=I(T_{hi}\geq s)$ and the maximum likelihood estimate of $h_0(s)$ is given in the equation (3.12). Differentiating with respect to $\boldsymbol{\gamma}$, we get

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h}\left\{\delta_{hi}E_i\boldsymbol{U_{hi}}-\sum_{k=1}^{K}h_o(s_k)E_i\{\boldsymbol{U_{hi}}\exp(\boldsymbol{x_{2i}}\boldsymbol{\beta_2}+\boldsymbol{\gamma}\boldsymbol{D_{hi}}\boldsymbol{U_{hi}})\}Z_{hi}(s_k)\right\}\tag{A.3}$$

where s_k is the baseline hazard at ordered failure times for $k = 1, 2, \ldots, K$.
As $h_0(s_k)$ is a function of γ, we will substitute $h_0(s_k)$ with its maximum likelihood estimate in (A.3). So,

$$\begin{aligned}\boldsymbol{S}_\gamma &= \sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\left\{E_i(\boldsymbol{U_{hi}}) - \frac{\sum_{k=1}^{K}E_i\{\boldsymbol{U_{hi}}\exp(\boldsymbol{x_{2i}\beta_2}+\boldsymbol{\gamma D_{hi}U_{hi}})\}Z_{hi}(s_k)}{\sum_{k=1}^{K}E_i\{\exp(\boldsymbol{x_{2i}\beta_2}+\boldsymbol{\gamma D_{hi}U_{hi}})\}Z_{hi}(s_k)}\right\} \\ &= \sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\left\{E_i(\boldsymbol{U_{hi}}) - \frac{E_i\{\boldsymbol{U_{hi}}\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}}{E_i\{\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}}\right\} \qquad \text{(A.4)}\end{aligned}$$

The derivative of $\boldsymbol{S}_\gamma$ function with respect to γ will give us the information. So,

$$\frac{\partial \boldsymbol{S}_\gamma}{\partial \boldsymbol{\gamma}} = -\sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\left\{\frac{E_i\{\boldsymbol{U_{hi}^2}\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}}{E_i\{\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}} - \frac{E_i^2\{\boldsymbol{U_{hi}}\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}}{E_i^2\{\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}}\right\} \qquad \text{(A.5)}$$

A.2 Gauss-Hermite Quadrature

EM algorithm requires some of the function of $E_i\{g(\boldsymbol{U_{hi}})\}$ in the maximization step. Gauss-Hermite quadrature is the main method that we need to apply in estimation of these functions. For example from equation (A.5), we need to calculate $E_i\{\exp(\boldsymbol{\gamma D_{hi}U_h i})\}$. As the distribution of $\mathbf{U}_{hi}|\mathbf{Y}_{hi}$ is $N(\mathbf{U}_{Y_{hi}}, \mathbf{W}_{Y_{hi}})$, $E_i\{\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\}$ can be evaluated as follows ignoring the constant terms

$$\begin{aligned}E_i\{\exp(\boldsymbol{\gamma D_{hi}U_{hi}})\} = \int\int_{-\infty}^{\infty}\exp\Big[&(\boldsymbol{\gamma D_{hi}U_{hi}}) + \delta_{hi}(\boldsymbol{\gamma D_{hi}U_{hi}}) \\ &- \int_0^{T_{hi}} h_0(s)\exp(\boldsymbol{\gamma D_{hi}U_{hi}})ds - (\boldsymbol{U_{hi}}-\boldsymbol{U_{Y_{hi}}})'\boldsymbol{W}_{\boldsymbol{Y_{hi}}}^{-1}(\boldsymbol{U_{hi}}-\boldsymbol{U_{Y_{hi}}})/2\Big]d\boldsymbol{U_{hi}} \qquad \text{(A.6)}\end{aligned}$$

The integral given in equation (A.6) is intractable. So, we need to apply the transformation of $\boldsymbol{U_{hi}}$ to $\boldsymbol{\eta_{hi}} = (2\boldsymbol{W}_{\boldsymbol{Y_{hi}}})^{-1/2}(\boldsymbol{U_{hi}} - \boldsymbol{U}_{\boldsymbol{Y_{hi}}})$ in order to use the Gauss-Hermite quadrature. $\boldsymbol{\eta_{hi}}$ has an $N(0,(1/2)I)$ distribution and the components are independent. In that case, with this transformation the $\boldsymbol{U_{hi}}$, say $\boldsymbol{U'_{hi}}$ become $\boldsymbol{U'_{hi}} = (\boldsymbol{2W}_{\boldsymbol{Y_{hi}}})^{1/2}\boldsymbol{\eta_{hi}} + \boldsymbol{U}_{\boldsymbol{Y_{hi}}}$. Thus, equation (A.6) becomes

$$\int\int_{-\infty}^{\infty}\exp\Big[(\gamma(u'_{h0i}+v'_{h0}))+\delta_{hi}(\gamma(u'_{h0i}+v'_{h0}))-\int_0^{T_{hi}}h_0(s)\exp(\gamma(u'_{h0i}+v'_{h0}))ds-(\eta_{h0i}^2+\eta_{h0}^2)\Big]d\boldsymbol{\eta_{hi}} \qquad \text{(A.7)}$$

By using the m-point Gauss-Hermite quadrature formula, this expression is as follows:

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\exp\Big[(\gamma(u'_{h0i}+v'_{h0}))+\delta_{hi}(\gamma(u'_{h0i}+v'_{h0}))-\int_0^{T_{hi}}h_0(s)\exp(\gamma(u'_{h0i}+v'_{h0}))ds\Big]w_j w_k \qquad \text{(A.8)}$$

where the components of $\boldsymbol{\eta_{hi}}$ takes m abscissa values and the corresponding weights are w_j and w_k.

A.3 Score Test results for Model D with separate association

	$Power_{V_1}$	$Power_{boot}$	Event Rate(%)	γ_1	γ_2	γ_3	γ_4
1	3	2	21.130	0	0	0	0
2	41	45	21.650	0.170	0	0	0
3	96	96	22.556	0.330	0	0	0
4	100	100	24.478	0.500	0	0	0
5	5	5	22.504	0	0.170	0	0
6	64	67	22.606	0.170	0.170	0	0
7	100	100	23.242	0.330	0.170	0	0
8	100	100	25.164	0.500	0.170	0	0
9	17	22	24.240	0	0.330	0	0
10	79	83	24.340	0.170	0.330	0	0
11	100	100	25.030	0.330	0.330	0	0
12	100	100	26.418	0.500	0.330	0	0
13	45	47	26.730	0	0.500	0	0
14	87	89	26.270	0.170	0.500	0	0
15	100	100	27.418	0.330	0.500	0	0
16	100	100	28.190	0.500	0.500	0	0
17	16	12	21.460	0	0	0.170	0
18	66	62	21.720	0.170	0	0.170	0
19	99	99	22.830	0.330	0	0.170	0
20	100	100	24.646	0.500	0	0.170	0
21	27	23	22.134	0	0.170	0.170	0
22	85	83	22.580	0.170	0.170	0.170	0
23	100	100	23.378	0.330	0.170	0.170	0
24	100	100	25.294	0.500	0.170	0.170	0
25	39	35	23.860	0	0.330	0.170	0
26	93	91	24.590	0.170	0.330	0.170	0
27	100	100	25.516	0.330	0.330	0.170	0
28	100	100	26.312	0.500	0.330	0.170	0
29	64	61	26.656	0	0.500	0.170	0
30	95	93	26.460	0.170	0.500	0.170	0
31	100	100	27.202	0.330	0.500	0.170	0
32	100	100	28.142	0.500	0.500	0.170	0
33	74	58	22.162	0	0	0.330	0
34	91	82	22.420	0.170	0	0.330	0
35	100	99	23.254	0.330	0	0.330	0

36	100	100	24.978	0.500	0	0.330	0
37	75	67	22.738	0	0.170	0.330	0
38	96	94	22.764	0.170	0.170	0.330	0
39	100	100	23.752	0.330	0.170	0.330	0
40	100	100	25.752	0.500	0.170	0.330	0
41	79	68	24.468	0	0.330	0.330	0
42	98	98	24.530	0.170	0.330	0.330	0
43	100	100	25.824	0.330	0.330	0.330	0
44	100	100	26.658	0.500	0.330	0.330	0
45	85	81	27.132	0	0.500	0.330	0
46	100	99	26.940	0.170	0.500	0.330	0
47	100	100	27.530	0.330	0.500	0.330	0
48	100	100	28.462	0.500	0.500	0.330	0
49	95	84	21.964	0	0	0.500	0
50	100	97	22.566	0.170	0	0.500	0
51	100	100	23.824	0.330	0	0.500	0
52	100	100	25.910	0.500	0	0.500	0
53	94	85	23.684	0	0.170	0.500	0
54	100	99	23.544	0.170	0.170	0.500	0
55	100	100	24.692	0.330	0.170	0.500	0
56	100	100	25.126	0.500	0.170	0.500	0
57	98	88	25.030	0	0.330	0.500	0
58	100	100	25.414	0.170	0.330	0.500	0
59	100	100	25.630	0.330	0.330	0.500	0
60	100	100	27.494	0.500	0.330	0.500	0
61	94	87	27.396	0	0.500	0.500	0
62	99	99	27.234	0.170	0.500	0.500	0
63	100	100	27.852	0.330	0.500	0.500	0
64	100	100	29.206	0.500	0.500	0.500	0
65	13	8	22.818	0	0	0	0.170
66	52	55	23.770	0.170	0	0	0.170
67	96	96	24.538	0.330	0	0	0.170
68	100	100	25.972	0.500	0	0	0.170
69	22	15	23.964	0	0.170	0	0.170
70	76	82	24.370	0.170	0.170	0	0.170
71	99	100	24.742	0.330	0.170	0	0.170
72	100	100	26.610	0.500	0.170	0	0.170
73	36	37	25.460	0	0.330	0	0.170
74	89	89	26.212	0.170	0.330	0	0.170
75	100	100	25.978	0.330	0.330	0	0.170
76	100	100	27.408	0.500	0.330	0	0.170
77	49	43	27.102	0	0.500	0	0.170
78	96	97	28.034	0.170	0.500	0	0.170
79	100	100	28.392	0.330	0.500	0	0.170
80	100	100	29.240	0.500	0.500	0	0.170
81	37	26	23.556	0	0	0.170	0.170

82	65	63	23.832	0.170	0	0.170	0.170
83	99	100	24.128	0.330	0	0.170	0.170
84	100	100	25.756	0.500	0	0.170	0.170
85	39	35	24.348	0	0.170	0.170	0.170
86	84	80	23.772	0.170	0.170	0.170	0.170
87	100	100	25.388	0.330	0.170	0.170	0.170
88	100	100	26.962	0.500	0.170	0.170	0.170
89	56	44	25.194	0	0.330	0.170	0.170
90	90	90	25.548	0.170	0.330	0.170	0.170
91	100	100	25.944	0.330	0.330	0.170	0.170
92	100	100	27.642	0.500	0.330	0.170	0.170
93	63	58	27.486	0	0.500	0.170	0.170
94	95	94	27.694	0.170	0.500	0.170	0.170
95	100	100	28.712	0.330	0.500	0.170	0.170
96	100	100	29.280	0.500	0.500	0.170	0.170
97	63	53	24.136	0	0	0.330	0.170
98	88	84	23.784	0.170	0	0.330	0.170
99	100	99	24.608	0.330	0	0.330	0.170
100	100	100	26.424	0.500	0	0.330	0.170
101	66	52	24.364	0	0.170	0.330	0.170
102	94	91	24.234	0.170	0.170	0.330	0.170
103	100	100	25.702	0.330	0.170	0.330	0.170
104	100	100	27.008	0.500	0.170	0.330	0.170
105	70	63	26.176	0	0.330	0.330	0.170
106	97	97	26.086	0.170	0.330	0.330	0.170
107	100	100	26.856	0.330	0.330	0.330	0.170
108	100	100	27.914	0.500	0.330	0.330	0.170
109	81	74	27.796	0	0.500	0.330	0.170
110	99	98	28.042	0.170	0.500	0.330	0.170
111	100	100	28.208	0.330	0.500	0.330	0.170
112	100	100	29.508	0.500	0.500	0.330	0.170
113	87	72	23.936	0	0	0.500	0.170
114	95	92	24.598	0.170	0	0.500	0.170
115	100	100	25.604	0.330	0	0.500	0.170
116	100	100	26.918	0.500	0	0.500	0.170
117	90	81	25.740	0	0.170	0.500	0.170
118	99	98	25.320	0.170	0.170	0.500	0.170
119	100	100	26.002	0.330	0.170	0.500	0.170
120	100	100	27.342	0.500	0.170	0.500	0.170
121	87	79	26.620	0	0.330	0.500	0.170
122	100	100	26.744	0.170	0.330	0.500	0.170
123	100	100	27.710	0.330	0.330	0.500	0.170
124	100	100	28.164	0.500	0.330	0.500	0.170
125	96	91	28.134	0	0.500	0.500	0.170
126	100	100	28.566	0.170	0.500	0.500	0.170
127	100	100	28.702	0.330	0.500	0.500	0.170

128	100	100	29.372	0.500	0.500	0.500	0.170
129	36	15	26.010	0	0	0	0.330
130	60	60	26.046	0.170	0	0	0.330
131	95	96	27.412	0.330	0	0	0.330
132	100	100	28.142	0.500	0	0	0.330
133	36	26	27.610	0	0.170	0	0.330
134	87	88	27.430	0.170	0.170	0	0.330
135	99	99	27.776	0.330	0.170	0	0.330
136	100	100	28.258	0.500	0.170	0	0.330
137	45	39	29.402	0	0.330	0	0.330
138	89	90	27.774	0.170	0.330	0	0.330
139	100	100	29.644	0.330	0.330	0	0.330
140	100	100	29.704	0.500	0.330	0	0.330
141	66	61	29.332	0	0.500	0	0.330
142	94	92	28.824	0.170	0.500	0	0.330
143	100	100	29.870	0.330	0.500	0	0.330
144	100	100	30.816	0.500	0.500	0	0.330
145	49	26	26.208	0	0	0.170	0.330
146	74	74	28.230	0.170	0	0.170	0.330
147	96	97	28.028	0.330	0	0.170	0.330
148	100	100	29.292	0.500	0	0.170	0.330
149	48	37	26.370	0	0.170	0.170	0.330
150	88	88	27.620	0.170	0.170	0.170	0.330
151	98	97	28.030	0.330	0.170	0.170	0.330
152	100	100	29.510	0.500	0.170	0.170	0.330
153	54	49	28.200	0	0.330	0.170	0.330
154	91	94	29.156	0.170	0.330	0.170	0.330
155	100	100	29.278	0.330	0.330	0.170	0.330
156	100	100	30.222	0.500	0.330	0.170	0.330
157	64	61	29.644	0	0.500	0.170	0.330
158	90	92	29.622	0.170	0.500	0.170	0.330
159	100	100	29.806	0.330	0.500	0.170	0.330
160	100	100	30.438	0.500	0.500	0.170	0.330
161	62	37	27.184	0	0	0.330	0.330
162	79	74	28.230	0.170	0	0.330	0.330
163	95	96	28.254	0.330	0	0.330	0.330
164	100	100	30.018	0.500	0	0.330	0.330
165	63	46	27.052	0	0.170	0.330	0.330
166	87	86	28.252	0.170	0.170	0.330	0.330
167	100	100	27.996	0.330	0.170	0.330	0.330
168	100	100	29.244	0.500	0.170	0.330	0.330
169	73	60	28.206	0	0.330	0.330	0.330
170	93	92	28.672	0.170	0.330	0.330	0.330
171	98	98	29.158	0.330	0.330	0.330	0.330
172	100	100	30.204	0.500	0.330	0.330	0.330
173	82	77	29.616	0	0.500	0.330	0.330

174	98	95	30.142	0.170	0.500	0.330	0.330
175	100	100	29.592	0.330	0.500	0.330	0.330
176	100	100	30.748	0.500	0.500	0.330	0.330
177	82	66	27.808	0	0	0.500	0.330
178	90	88	28.360	0.170	0	0.500	0.330
179	100	100	29.430	0.330	0	0.500	0.330
180	100	100	29.200	0.500	0	0.500	0.330
181	78	65	28.242	0	0.170	0.500	0.330
182	94	92	28.144	0.170	0.170	0.500	0.330
183	100	100	29.554	0.330	0.170	0.500	0.330
184	100	100	29.928	0.500	0.170	0.500	0.330
185	85	73	29.472	0	0.330	0.500	0.330
186	98	96	29.240	0.170	0.330	0.500	0.330
187	100	100	29.688	0.330	0.330	0.500	0.330
188	100	100	29.632	0.500	0.330	0.500	0.330
189	83	80	29.950	0	0.500	0.500	0.330
190	100	100	30.258	0.170	0.500	0.500	0.330
191	100	100	30.386	0.330	0.500	0.500	0.330
192	100	100	30.834	0.500	0.500	0.500	0.330
193	42	25	30.264	0	0	0	0.500
194	60	59	30.856	0.170	0	0	0.500
195	90	95	31.480	0.330	0	0	0.500
196	99	99	32.214	0.500	0	0	0.500
197	46	30	30.508	0	0.170	0	0.500
198	77	77	30.584	0.170	0.170	0	0.500
199	95	97	30.176	0.330	0.170	0	0.500
200	100	100	31.676	0.500	0.170	0	0.500
201	51	36	30.308	0	0.330	0	0.500
202	79	83	30.896	0.170	0.330	0	0.500
203	96	99	30.754	0.330	0.330	0	0.500
204	100	99	31.652	0.500	0.330	0	0.500
205	65	58	32.590	0	0.500	0	0.500
206	88	91	31.828	0.170	0.500	0	0.500
207	100	99	32.386	0.330	0.500	0	0.500
208	100	100	33.160	0.500	0.500	0	0.500
209	56	35	30.700	0	0	0.170	0.500
210	62	63	30.968	0.170	0	0.170	0.500
211	95	95	30.656	0.330	0	0.170	0.500
212	100	100	31.606	0.500	0	0.170	0.500
213	63	40	30.540	0	0.170	0.170	0.500
214	71	66	29.522	0.170	0.170	0.170	0.500
215	98	100	31.180	0.330	0.170	0.170	0.500
216	100	100	32.994	0.500	0.170	0.170	0.500
217	57	48	30.984	0	0.330	0.170	0.500
218	89	90	31.426	0.170	0.330	0.170	0.500
219	98	99	31.796	0.330	0.330	0.170	0.500

220	100	100	32.104	0.500	0.330	0.170	0.500
221	64	60	32.370	0	0.500	0.170	0.500
222	89	92	31.644	0.170	0.500	0.170	0.500
223	100	100	31.736	0.330	0.500	0.170	0.500
224	100	100	33.174	0.500	0.500	0.170	0.500
225	61	38	29.906	0	0	0.330	0.500
226	78	69	30.834	0.170	0	0.330	0.500
227	96	96	31.140	0.330	0	0.330	0.500
228	99	100	31.410	0.500	0	0.330	0.500
229	62	44	30.828	0	0.170	0.330	0.500
230	87	90	30.760	0.170	0.170	0.330	0.500
231	99	100	30.638	0.330	0.170	0.330	0.500
232	100	100	31.704	0.500	0.170	0.330	0.500
233	65	55	31.070	0	0.330	0.330	0.500
234	87	92	30.266	0.170	0.330	0.330	0.500
235	99	99	32.066	0.330	0.330	0.330	0.500
236	100	100	31.560	0.500	0.330	0.330	0.500
237	71	64	31.260	0	0.500	0.330	0.500
238	97	97	32.478	0.170	0.500	0.330	0.500
239	100	100	32.182	0.330	0.500	0.330	0.500
240	100	100	33.274	0.500	0.500	0.330	0.500
241	72	54	30.798	0	0	0.500	0.500
242	79	77	31.562	0.170	0	0.500	0.500
243	100	99	31.624	0.330	0	0.500	0.500
244	100	100	32.426	0.500	0	0.500	0.500
245	73	57	30.274	0	0.170	0.500	0.500
246	84	91	31.112	0.170	0.170	0.500	0.500
247	100	100	32.130	0.330	0.170	0.500	0.500
248	100	100	32.596	0.500	0.170	0.500	0.500
249	77	65	31.006	0	0.330	0.500	0.500
250	96	99	31.750	0.170	0.330	0.500	0.500
251	99	99	31.586	0.330	0.330	0.500	0.500
252	100	100	32.398	0.500	0.330	0.500	0.500
253	89	84	32.254	0	0.500	0.500	0.500
254	95	94	33.148	0.170	0.500	0.500	0.500
255	99	100	32.860	0.330	0.500	0.500	0.500
256	100	100	32.658	0.500	0.500	0.500	0.500

Table A.1: The powers of the score test, event rates and association parameters for each random effect component based on 100 simulation

A.4 `Simjointml()` function

```
simjointml<- function(m, noh, beta, beta2, sig.v, sig.hv, sig.err, gamma,
 ntms=8, theta0=-3,censlam=exp(-3), censoring=T, truncation=T){
```

```
trunctime=max(ntms)
 n<- m*noh
vi<-rnorm(n,0,sig.v)#vi<-mvrnorm(n ,mu=c(0,0),Sigma=var.v)
vih<-rnorm(noh, 0, sig.hv)
ctsx<-rnorm(n)
grp<-rbinom(n,1,0.5)# binary covariate for treatment
err<-rnorm(n*ntms, 0, sig.err)
t<-c(0:(ntms-1)) #time covariate
y<- rep(0, n*ntms)
X<- cbind(1,rep(t,n),rep(grp, each=ntms),rep(ctsx,each=ntms))#create
# design matrix
Xbeta<- X%*%beta#fixed effects
b<- cbind(rep(vi,each=ntms))
#b<- cbind(rep(vi[,1],each=10), rep(vi[,2],each=10))
h1<-cbind(rep(vih, each=ntms*m))
Y<- Xbeta+(b+h1)+err #generate Y
##### generate data for coxph
X2 <- cbind(ctsx,grp)
Xbeta2 <-X2%*%beta2
svih<-rep(vih,each=m)
uu<-runif(n)
cens<-rep(1,n)
survtime<--log(uu)/exp(theta0+Xbeta2+gamma[1]*vi+gamma[2]*svih)
censtime <- -log(runif(n))/censlam
censtime <- pmin(censtime,trunctime)
ii<-censtime<survtime
survtime[ii]<-censtime[ii]
cens[ii]<-0



idl<-rep(1:n,each=ntms)
hosidl<-rep(1:noh, each=ntms*m)
```

```
  obstime<-rep(0:(ntms-1),n)
  lsurvtime<-rep(survtime, each=ntms)
  l.cens<-rep(cens, each=ntms)
  ctsxl<-rep(ctsx,each=ntms)
  grpl<-rep(grp,each=ntms)
  ul<-rep(vi, each=ntms)
  uhl<-rep(vih, each=m*ntms)
  ###put all the data into a dataframe

  obstimecopy<-rep(0:(ntms-1),n)
  Y<- Y[obstime<=lsurvtime]
  idl<- idl[obstime<=lsurvtime]
  hosidl<-hosidl[obstime<=lsurvtime]
  l.cens<-l.cens[obstime<=lsurvtime]
  ctsxl<-ctsxl[obstime<=lsurvtime]
  grpl<- grpl[obstime<=lsurvtime]
  ul<-ul[obstime<=lsurvtime]
  uhl<-uhl[obstime<=lsurvtime]
  obsertime<-obstimecopy[obstime<=lsurvtime]
  lsurvtime<-lsurvtime[obstime<=lsurvtime]
  cat(100*sum(cens)/n,"% experienced event\n")
  event.rate <- 100*sum(cens)/n
  fulldata<-data.frame(y=Y,IDL=idl ,hospitalid=hosidl,
                       survtime=lsurvtime, obntime=obsertime, ctsxl= ctsxl,
                       group=grpl, censoring=l.cens, u0=ul, u1=uhl)
  longdat<- fulldata
 # survdat<-  longdat[!duplicated(longdat$IDL),]
  list(longdat=longdat, eventrate=event.rate)
}
```

Bibliography

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120.

Andersen, P. K., Gill, R. D., Borgan, Ø., and Keiding, N. (1993). Statistical models based on counting processes.

Antolini, L., Boracchi, P., and Biganzoli, E. (2005). A time-dependent discrimination index for survival data. *Statistics in Medicine*, 24(24):3927–3944.

Asar, Ö., Ritchie, J., Kalra, P. A., and Diggle, P. J. (2015). Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International journal of epidemiology*, 44(1):334–344.

Austin, P. C. (2012). Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in medicine*, 31(29):3946–3958.

BAŞAR, E. (2017). Aalen's additive, cox proportional hazards and the cox-aalen model: Application to kidney transplant data. *Sains Malaysiana*, 46(3):469–476.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, 24(11):1713–1723.

Berzuini, C. and Larizza, C. (1996). A unified approach for modeling longitudinal and failure time data, with application in medical monitoring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):109–123.

Borucka, J. (2014). Methods for handling tied events in the cox proportional hazard model. *Studia Oeconomica Posnaniensia*, 2(2):91–106.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, 1(1):49–67.

Chi, Y.-Y. and Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445.

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

Colosimo, E., Ferreira, F. V., Oliveira, M., and Sousa, C. (2002). Empirical comparisons between kaplan-meier and nelson-aalen survival function estimators. *Journal of Statistical Computation and Simulation*, 72(4):299–308.

Cox, D. R. (1972). Regression models and life-tables. In *Breakthroughs in Statistics*, pages 527–541. Springer.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Crowder, M. (2017). *Analysis of repeated measures*. Routledge.

Crowther, M. J. (2014). *Development and application of methodology for the parametric analysis of complex survival and joint longitudinal-survival data in biomedical research*. PhD thesis, Department of Health Sciences.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Diggle, P. (1998). Dealing with missing values in longitudinal studies. *Statistical Analysis of Medical Data*, pages 203–228.

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford University Press.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics*, pages 959–971.

Elashoff, R., Li, G., and Li, N. (2016). *Joint Modeling of Longitudinal and Time-to-event Data*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. CRC Press LLC.

Eloranta, S., Lambert, P. C., Andersson, T. M., Czene, K., Hall, P., Björkholm, M., and Dickman, P. W. (2012). Partitioning of excess mortality in population-based cancer patient survival studies using flexible parametric survival models. *BMC medical research methodology*, 12(1):86.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685.

Fennema-Notestine, C., Hagler, D. J., McEvoy, L. K., Fleisher, A. S., Wu, E. H., Karow, D. S., and Dale, A. M. (2009). Structural mri biomarkers for preclinical and mild alzheimer's disease. *Human brain mapping*, 30(10):3238–3253.

Fieuws, S., Verbeke, G., Maes, B., and Vanrenterghem, Y. (2007). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, 9(3):419–431.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal data analysis*. CRC Press.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis*, volume 998. John Wiley & Sons.

Gerds, T. A. and Schumacher, M. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040.

Gueorguieva, R. and Sanacora, G. (2006). Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine*, 25(8):1307–1322.

Guo, X. and Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090.

Heagerty, P. J. and Zheng, Y. (2005). Survival model predictive accuracy and roc curves. *Biometrics*, 61(1):92–105.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.

Henderson, R., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics*, 3(1):33–50.

Hickey, G., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2018). A joint model and software package for time-to-event and multivariate longitudinal data.

Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology*, 16(1):117.

Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2017). *joineRML: Joint Modelling of Multivariate Longitudinal Data and Time-to-Event Outcomes*. R package version 0.4.0.

Hosmer, D. W., Lemeshow, S., and May, S. (2008). Model development. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition*, pages 132–168.

Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4):1037–1043.

Huang, X., Stefanski, L. A., and Davidian, M. (2009). Latent-model robustness in joint models for a primary endpoint and a longitudinal process. *Biometrics*, 65(3):719–727.

Ibrahim, J. G. and Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test*, 18(1):1–43.

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C., et al. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of magnetic resonance imaging*, 27(4):685–691.

Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J. M., and Commenges, D. (2010). Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics*, 66(1):11–19.

Jaffa, M. A., Woolson, R. F., and Lipsitz, S. R. (2011). Slope estimation for bivariate longitudinal outcomes adjusting for informative right censoring by using a discrete survival model: application to the renal transplant cohort. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):387–402.

Jagust, W. J., Bandy, D., Chen, K., Foster, N. L., Landau, S. M., Mathis, C. A., Price, J. C., Reiman, E. M., Skovronsky, D., and Koeppe, R. A. (2010). The alzheimer's disease neuroimaging initiative positron emission tomography core. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 6(3):221–229.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, pages 805–820.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

King, N. B., Harper, S., and Young, M. E. (2012). Use of relative and absolute effect measures in reporting health inequalities: structured review. *Bmj*, 345:e5774.

Kleinbaum, D. G. (1998). Survival analysis, a self-learning text. *Biometrical Journal*, 40(1):107–108.

Ko, F.-S. (2010). Using frailty models to identify the longitudinal biomarkers in survival analysis. *Communications in Statistics—Theory and Methods*, 39(18):3222–3237.

Ko, F.-S. (2014a). Identification of longitudinal biomarkers for survival by a score test derived from a joint model of longitudinal and competing risks data. *Journal of Applied Statistics*, 41(10):2270–2281.

Ko, F.-S. (2014b). Identification of longitudinal biomarkers in survival analysis for competing risks data. *Communications in Statistics-Theory and Methods*, 43(16):3329–3342.

Ko, F.-S. (2014c). The method to identify a biomarker and to evaluate its efficiency for survival by using the joint model of the accelerate failure time and longitudinal data. *Communications in Statistics-Theory and Methods*, 43(1):72–89.

Ko, F.-S. (2016). Identification of potential longitudinal biomarkers under the accelerated failure time model in multivariate survival data. *Communications in Statistics-Theory and Methods*, 45(3):655–669.

Ko, F.-S. (2017). Application of trajectories from growth curve in identification of longitudinal biomarker for the multivariate survival data. *Journal of Applied Statistics*, 44(3):416–426.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Landau, S. M., Harvey, D., Madison, C. M., Koeppe, R. A., Reiman, E. M., Foster, N. L., Weiner, M. W., and Jagust, W. J. (2011). Associations between cognitive, functional, and fdg-pet measures of decline in ad and mci. *Neurobiology of aging*, 32(7):1207–1218.

Landau, S. M., Mintun, M. A., Joshi, A. D., Koeppe, R. A., Petersen, R. C., Aisen, P. S., Weiner, M. W., and Jagust, W. J. (2012). Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of neurology*, 72(4):578–586.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., and Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the dia bayesian joint modeling working group. *Statistics in medicine*, 34(14):2181–2195.

Lee, E. T. and Wang, J. (2003). *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons.

Li, K., Chan, W., Doody, R. S., Quinn, J., and Luo, S. (2017). Prediction of conversion to alzheimer's disease with longitudinal measures and time-to-event data. *Journal of Alzheimer's Disease*, (Preprint):1–11.

Lillard, L. A. and Panis, C. W. (2003). aml multilevel multiprocess statistical software, version 2.0. *EconWare, Los Angeles, California*.

Lin, H., McCulloch, C. E., and Mayne, S. T. (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21(16):2369–2382.

Lin, X., Taylor, J. M., and Ye, W. (2008). A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and its Interface*, 1(1):33–45.

Lindstrom, M. J. and Bates, D. M. (1988). Newton—raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404):1014–1022.

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134.

Little, R. J. (2008). Selection and pattern-mixture models. *Longitudinal data analysis*, pages 409–431.

Little, R. J. and Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.

Liu, L., Ma, J. Z., and O'Quigley, J. (2008). Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (esrd) data. *Statistics in medicine*, 27(27):5679–5691.

McCrink, L. M., Marshall, A. H., and Cairns, K. J. (2013). Advances in joint modelling: A review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81(2):249–269.

McCulloch, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research*, 17(1):53–73.

Meng, X.-L. and Van Dyk, D. (1997). The em algorithm—an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):511–567.

Molenberghs, G. and Kenward, M. (2007). *Missing data in clinical studies*, volume 61. John Wiley & Sons.

Mood, A., Graybill, F., and Boes, D. (1974). Introduction to statistical theory.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W., and Beckett, L. (2005a). The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4):869–877.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., Trojanowski, J. Q., Toga, A. W., and Beckett, L. (2005b). Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). *Alzheimer's & Dementia*, 1(1):55–66.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966.

Oakes, D. (2013). An introduction to survival models: in honor of ross prentice. *Lifetime data analysis*, 19(4):442–462.

Pantazis, N. and Touloumi, G. (2007). Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: A simulation study. *Statistics in medicine*, 26(30):5473–5485.

Pantazis, N., Touloumi, G., Walker, A., and Babiker, A. (2005). Bivariate modelling of longitudinal measurements of two human immunodeficiency type 1 disease progression markers in the presence of informative drop-outs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(2):405–423.

Pencina, M. J., D'Agostino Sr, R. B., D'Agostino Jr, R. B., and Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine*, 27(2):157–172.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology*, 56(3):303–308.

Philipson, P., Sousa, I., Diggle, P. J., Williamson, P., Kolamunnage-Dona, R., Henderson, R., and Hickey, G. L. (2017). *joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data*. R package version 1.2.2.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2017). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-131.

Pinheiro, J. C., Liu, C., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10(2):249–276.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). Numerical recipes in fortran (cambridge.

Proust-Lima, C., Séne, M., Taylor, J. M., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Statistical methods in medical research*, 23(1):74–90.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratcliffe, S. J., Guo, W., and Ten Have, T. R. (2004). Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*, 60(4):892–899.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press.

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software*, 72(7):1–45.

Rizopoulos, D. and Ghosh, P. (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12):1366–1380.

Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential laplace approximations for the joint modelling of survival and longitudinal data.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71(3):637–654.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for non-ignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120.

Schemper, M. and Henderson, R. (2000). Predictive accuracy and explained variation in cox regression. *Biometrics*, 56(1):249–255.

Song, X., Davidian, M., and Tsiatis, A. A. (2002a). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, 3(4):511–528.

Song, X., Davidian, M., and Tsiatis, A. A. (2002b). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753.

Sousa, I. (2011). A review on joint modelling of longitudinal measurements and time-to-event. *REVSTAT–Statistical Journal*, 9(1):57–81.

Sweeting, M. J. and Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763.

Tang, A.-M. and Tang, N.-S. (2015). Semiparametric bayesian inference on skew–normal joint modeling of multivariate longitudinal and survival data. *Statistics in medicine*, 34(5):824–843.

Tashkin, D. P., Elashoff, R., Clements, P. J., Goldin, J., Roth, M. D., Furst, D. E., Arriola, E., Silver, R., Strange, C., Bolster, M., et al. (2006). Cyclophosphamide versus placebo in scleroderma lung disease. *New England Journal of Medicine*, 354(25):2655–2666.

Taylor, J. M., Cumberland, W., and Sy, J. (1994). A stochastic model for analysis of longitudinal aids data. *Journal of the American Statistical Association*, 89(427):727–736.

Therneau, T. M. (2015). *A Package for Survival Analysis in S*. version 2.38.

Tsiatis, A., Degruttola, V., and Wulfsohn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statistics in medicine*, 19(24):3309–3324.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Verbeke, G. (1997). Linear mixed models for longitudinal data. In *Linear mixed models in practice*, pages 63–153. Springer.

Williamson, P., Kolamunnage-Dona, R., Philipson, P., and Marson, A. (2008). Joint modelling of longitudinal and competing risks data. *Statistics in medicine*, 27(30):6426–6438.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):375–387.

Yamaguchi, T., Ohashi, Y., and Matsuyama, Y. (2002). Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistical methods in medical research*, 11(3):221–236.

Yu, M., Law, N. J., Taylor, J. M., and Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, pages 835–862.